



**Maria Madalena
Gomes Fonseca**

**Modelo Bayesiano do Aluno no Cálculo com Várias
Variáveis**



**Maria Madalena
Gomes Fonseca**

**Modelo Bayesiano do Aluno no Cálculo com Várias
Variáveis**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Matemática e Aplicações, Área de especialização em Estatística e Investigação Operacional, realizada sob a orientação científica do Doutor Luís António Arsénio Descalço, Professor Auxiliar do Departamento de Matemática da Universidade de Aveiro e da Doutora Maria Paula Lopes dos Reis Carvalho, Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro.

o júri / the jury

presidente / president

Professora Doutora Isabel Maria Simões Pereira

Professor Auxiliar da Universidade de Aveiro

vogais / examiners committee

**Professora Doutora Maria Teresa Mesquita Cunha Machado
Malheiro**

Professora Auxiliar da Universidade do Minho - Escola de Ciências

Professor Doutor Luís António Arsénio Descalço

Professor Auxiliar da Universidade de Aveiro (orientador)

agradecimentos / acknowledgements

Começo por agradecer aos meus orientadores, o professor Dr. Luís Descalço e a professora Dra. Paula Carvalho, sem os quais este trabalho não teria sido possível. Agradeço-lhes a sua dedicação, paciência, compreensão, os seus ensinamentos, as suas sugestões e a sua ajuda. À minha amiga Olga Oliveira, por ser a responsável pela minha decisão em fazer este percurso e pelo seu apoio, disponibilidade e incentivo constantes ao longo dele. A minha eterna gratidão pois sem ela seria difícil levar isto a bom porto.

Aos meus filhos, Marta e Manuel, pela compreensão, sorrisos e ternura que sempre demonstraram, apesar da falta de atenção a que, por vezes, estiveram sujeitos.

Obrigada aos meus pais e à minha irmã pelo amor incondicional e pela forma como ao longo de todos estes anos me apoiaram e valorizaram. Queria agradecer, ainda, ao meu marido, António Lemos, por acreditar em mim e me proporcionar esta oportunidade, apesar do encargo acrescido, a vários níveis, a que foi sujeito ao longo deste percurso.

Por fim, agradeço a todos que de alguma forma me ajudaram a concluir este trabalho.

Palavras-chave

STI, Probabilidades, Teorema de Bayes, Grafos, Redes bayesianas, Modelo de aluno

Resumo

O trabalho apresentado nesta dissertação consiste no estudo de uma aplicação de redes bayesianas num modelo do utilizador, que foi implementado numa aplicação web de apoio ao ensino. Esta aplicação é denominada SIACUA - Sistema Interativo de Aprendizagem por Computador, Universidade de Aveiro e tem como suporte gerador de questões os projetos PmatE, Megua. A utilização deste sistema é feita via internet em <http://siacua.web.ua.pt/>. Apresenta-se um exemplo de aplicação para a avaliação dos conhecimentos dos alunos na disciplina de Cálculo III, ilustrando a funcionalidade do sistema. Este transmite informações, em termos probabilísticos, ao aluno, sobre quais os temas do programa da disciplina que domina. Faz-se, também, neste trabalho uma introdução aos modelos probabilísticos, teoria de grafos e redes bayesianas.

Palavras-chave

STI, Probability, Bayes' theorem, Graphs, Bayesian Networks, Student Model

Resumo

This thesis presents a user's model approach of a Bayesian network application, which is implemented in a web application and used as an aid to support education. This application called SIACUA, which is a Portuguese acronym for the University of Aveiro Computer Interactive Learning System, has questions generated from the PmatE, Megua projects. This system works on the <http://siacua.web.ua.pt/> site. An example of the functionality of the application is shown through the evaluation of the student's knowledge in the subject of Calculus III. It shows the student, in probabilistic terms, what contents of the subject's curriculum that are mastered. It is also makes an introduction to probabilistic models, graph theory and Bayesian networks.

Conteúdo

Conteúdo	i
Lista de Figuras	iii
Lista de Tabelas	vii
1 Introdução	1
1.1 Sistema Tutor Inteligentes STI	2
1.2 Modelo de aluno em STI	3
1.3 Objetivos gerais do trabalho	4
1.4 Organização da dissertação	5
2 Grafos	7
3 Probabilidades	19
3.1 Conceitos Iniciais	20
3.2 Cálculo de Probabilidades. Lei de Laplace	25
3.3 Probabilidade Condicionada	27
3.4 Acontecimentos Independentes	30
3.5 Teorema da probabilidade total e Regra de Bayes	31
3.6 Distribuição de Probabilidade	35
4 Redes Bayesianas	43

4.1	Apresentação Intuitiva	45
4.2	Apresentação Formal	53
5	Modelo genérico do aluno baseado em redes bayesianas	59
5.1	Variáveis definição	60
5.2	Definição da estrutura	61
5.3	Estimativa de parâmetros	62
5.3.1	Estimativa de parâmetros para nós que representam conceitos básicos	62
5.3.2	Estimativa de parâmetros para nós que representam temas	63
5.3.3	Estimativa de parâmetros para nós que representam evidências . . .	66
5.3.4	Simplificação de parâmetros a especificar para nós que representam evidências (provas)	70
6	A aplicação web SIACUA	77
6.1	Apresentação da aplicação	79
6.2	Mapa conceptual no domínio de Cálculo III	82
6.3	Modelação da rede bayesiana	85
6.3.1	Na rede de conceitos	85
6.3.2	Na rede de conceitos-evidência	88
6.4	Questões	97
6.4.1	PmatE	97
6.4.2	MEGUA	99
6.4.3	Dinâmica dos três sistemas - MEGUA, PmatE e SIACUA	100
6.5	Implementação de Software	102
6.6	Experiências/ Resultados	102
7	Conclusão	105
	Referências	107

Lista de Figuras

2.1	Grafo $G(V, E)$, com $V = \{A, B, C, D, F\}$ e $E = \{e_1, e_2, e_3, e_4, e_5\}$	8
2.2	Multigrafo. AB e BA são arestas paralelas e CC é um lacete.	8
2.3	Digrafo $G(V, \vec{A})$, com $V = \{A, B, C, D, F\}$ e $\vec{A} = \{e_1, e_2, e_3, e_4, e_5\}$	9
2.4	Supergrafo e subgrafo.	10
2.5	Subgrafo induzido.	10
2.6	Grafo K_4	11
2.7	Grafo N_4	11
2.8	União e interseção de grafos.	12
2.9	Grafo G do exemplo 2.13.	13
2.10	Grafo G do exemplo 2.14.	14
2.11	Uma árvore.	14
2.12	Uma floresta.	15
2.13	Árvore abrangente do Grafo G	16
2.14	Digrafo do exemplo 2.19.	17
2.15	Digrafo do exemplo 2.19.	18
3.1	Partição de Ω	32
3.2	Representação do teorema da probabilidade total.	33
3.3	Representação esquemática de uma variável aleatória.	35
4.1	Conexão serial.	44
4.2	Conexão divergente.	44

4.3	Conexão convergente.	44
4.4	Exemplo de uma rede bayesiana.	46
4.5	Exemplo de uma rede bayesiana em medicina.	46
4.6	Rede bayesiana completa.	47
4.7	Rede bayesiana do exemplo 4.4.	51
4.8	Rede bayesiana com duas variáveis.	54
4.9	Rede bayesiana com cinco variáveis.	55
4.10	Rede bayesiana com três variáveis.	56
5.1	Uma estrutura do modelo de aluno com nós conhecimento.	61
5.2	A estrutura do MGARB.	62
5.3	Excerto da rede da figura 5.2.	63
5.4	Rede bayesiana com variáveis conhecimento.	64
5.5	Rede bayesiana de conceitos.	65
5.6	Tabela de probabilidades condicionada do exemplo 5.3.	66
5.7	Rede bayesiana com evidência.	67
5.8	Rede bayesiana - propagação.	69
5.9	Tabela de probabilidades condicionadas associada ao nó da resposta R . . .	69
5.10	Relações entre conceito – pergunta tipo teste (Itens).	70
5.11	Gráfico da função $G(x)$ usada para calcular as probabilidades condicionadas. .	71
5.12	Uso do $G(x)$ para determinar as probabilidades.	72
5.13	Rede bayesiana com evidência.	74
6.1	Página inicial da aplicação SIACUA.	79
6.2	Página da SIACUA relativa ao Cálculo III.	80
6.3	Progressos na SIACUA.	80
6.4	Escolha das questões na SIACUA.	81
6.5	Mapa conceptual: rede bayesiana com nós que representam conceitos. . . .	83
6.6	Excerto do mapa conceptual.	84
6.7	Excerto de uma rede com as probabilidades marginais.	86

6.8	Tabela de probabilidades condicionadas do nó C_{1310}	87
6.9	Tabela de probabilidades condicionadas do nó C_{1320}	87
6.10	Excerto da rede com os respetivos pesos.	88
6.11	Excerto da rede com adição do nó Q do exemplo 6.3.	89
6.12	Tabela de probabilidades condicionadas do nó Q , obtidas no modelo. . . .	91
6.13	Excerto da rede com adição do nó Q do exemplo 6.4.	91
6.14	Probabilidades condicionadas de Q	92
6.15	Atualização do estado de conhecimento dos nós C_{1310} , C_{1300} e C_{1000}	93
6.16	Probabilidades dos nós do exemplo 6.5.	94
6.17	Probabilidades obtidas no GeNIe para C_{1311} e C_{1312}	95
6.18	Probabilidades obtidas no GeNIe para C_{1310}	95
6.19	Probabilidades obtidas no GeNIe para C_{1300}	96
6.20	Barras de progresso do aluno.	96
6.21	Página do PmatE.	97
6.22	Uma questão de PmatE.	99
6.23	Página do MEGUA.	99
6.24	Uma questão de MEGUA.	101
6.25	Resultados do inquérito.	103

Lista de Tabelas

3.1	Propriedades das operações entre acontecimentos.	23
3.2	Distribuição de probabilidade de X	38
3.3	Tabela da função de probabilidade conjunta do exemplo 3.18.	39
3.4	Tabela da função de probabilidade conjunta completa do exemplo 3.18. . .	40
4.1	Tabela de probabilidades conjuntas.	48
4.2	Tabela de probabilidades Condicionadas.	48
4.3	Tabela de probabilidades conjuntas.	51
5.1	Distribuições condicionadas do modelo MGARB.	64
6.1	Tabela de probabilidades condicionadas do nó Q , obtidas analiticamente. .	90
6.2	Cálculo das probabilidades condicionadas usando a função G	92
6.3	Probabilidades marginais dos nós C_{1311} e C_{1312} obtidas analiticamente. . .	94

Capítulo 1

Introdução

O ensino de matemática, não é tarefa fácil, na medida em que esta disciplina é vista desde sempre e por uma grande parte dos alunos, como difícil de aprender. Prova disso são os níveis de insucesso na disciplina em comparação com outras em todos os níveis de escolaridade. Uma das grandes preocupações dos professores é encontrar estratégias para incentivar, nos alunos, o gosto em aprender matemática e terem sucesso na disciplina. Assim, novas práticas e metodologias têm sido aplicadas para motivar os alunos e permitir uma aprendizagem mais eficaz. A par da rápida evolução da tecnologia e da consequente conceção do ensino, foram desenvolvidos diversos mecanismos e ferramentas para implementação e suporte na educação. É neste contexto que, por exemplo, o projeto PmatE promove todos os anos várias iniciativas que pretendem envolver o número máximo de escolas, professores e alunos de todo o país [20].

A utilização de computadores, para melhorar o desempenho e a motivação dos alunos é reconhecido no relatório *Mathematical Advisory Panel*¹ nos EUA, onde se destacam os efeitos positivos no desempenho dos alunos em matemática com a utilização de *softwares* de instrução em comparação com a não incorporação dessas tecnologias [4].

As crianças cada vez mais cedo estão expostas à comunicação e informação tecnológica, o que as torna especialistas neste domínio, com uma enorme quantidade de páginas da

¹The Final Report of the National Mathematics Advisory Panel. U.S. Department of Education (2008)

web, vídeos, e outros materiais de aprendizagem tão facilmente acessíveis. Para os alunos de hoje, o processo de aprendizagem é substancialmente diferente do que era na geração anterior, que inclui os seus pais e professores. O método tradicional de ensino baseado na compreensão estática de textos escritos e na transmissão teórica e exaustiva do conhecimento, está ultrapassado. O aluno passou a ter um papel ativo no processo de ensino aprendizagem, num contexto em que as habilidades do estudante são desafiadas, exigindo a sua participação ativa, sendo os mesmos capazes de controlar e avaliar a sua aprendizagem. Isto não significa que o professor não continue a ter o papel mais importante, para que isso seja concretizado. Cabe-lhe a ele, além das aulas presenciais e da avaliação escrita, o desenvolvimento de ambientes de boa aprendizagem, elaborando e disponibilizando material didático, permitindo aos alunos aplicar os conceitos, resolver problemas, discutir ideias e ter um feedback imediato sobre o seu progresso e o desenvolvimento das suas competências[2].

1.1 Sistema Tutor Inteligentes STI

Com a universalização do ensino, a diversidade de alunos, a nível de capacidades, existente nas escolas é cada vez maior, bem como o elevado número de alunos por turma. A necessidade de criar um novo tipo de sistemas de ensino assistido por computador que se centre em cada aluno como ser individual, isto é, que seja capaz de se adaptar a cada aluno como um ser singular, são os chamados Sistemas Tutores Inteligentes (STI). Estes novos sistemas de ensino assistido por computador armazenam as informações de cada aluno, através da interação deste com o sistema, criando, assim, um comportamento padrão de cada aluno. Estes sistemas usam depois esta informação para se adaptarem ao aluno, bem como, para lhe transmitir algum feedback sobre a sua aprendizagem, adaptando-se portanto, ao desempenho de cada um. Os STIs são programas que contêm um amplo conhecimento sobre um determinado assunto, e cujo objetivo é transmitir esse conhecimento aos alunos através de um processo interativo individualizado, tentando imitar a forma como um professor iria orientar o aluno no seu processo de aprendizagem [17]. O uso de

STIs, neste contexto, é de grande utilidade, na medida em que cada aluno, fora da aula, pode avançar consoante o seu ritmo de estudo e de aprendizagem, tendo, também, apoio e material para esse fim [4], [2].

1.2 Modelo de aluno em STI

A adaptabilidade ao aluno, nos sistemas tutores inteligentes não é tarefa fácil e é alvo de várias pesquisas, pois tem de ter em conta as várias características do aluno. Dessa forma, o modelo do aluno é de importância fundamental, pois tem a capacidade de se adaptar a cada aluno, usando para isso as informações individuais de cada um, que vai armazenando. É nesse ponto que as técnicas de Inteligência Artificial (IA) têm sido utilizadas, na medida em toma decisões tendo em conta a informação armazenada. O modelo de aluno tem por objetivo proporcionar apoio interativo ao aluno sendo, por isso, necessário que o ambiente recolha informações sobre o conhecimento do aluno. Considerando-se que o nível de conhecimento do aluno é uma informação imprecisa, essa modelação envolve bastante incerteza, até porque, o aluno pode seguir vários raciocínios que não precisa demonstrar de forma explícita no processo de interação. Até mesmo os professores lidam com essa incerteza, optando-se, assim, pelo uso de redes bayesianas (RBs), também chamadas de redes de crença. As redes bayesianas são uma das principais técnicas usadas na construção de modelos de aluno, sendo consideradas adequadas para ambientes em que existe incerteza, uma vez que empregam o raciocínio probabilístico, o qual permite identificar um certo grau de crença sobre o nível de conhecimento do aluno. Por isso, as RBs têm sido empregues em STIs nas inferências sobre o comportamento do aluno e nas tomadas de decisões sobre as ações do professor.

Diante disso, foi proposto e implementado o emprego de um modelo bayesiano do aluno de um sistema tutor inteligente (STI) para auxiliar no processo de aprendizagem de conceitos de várias disciplinas de matemática na universidade de Aveiro. A escolha deste modelo deve-se ao facto das probabilidades condicionadas necessárias serem automaticamente definidas a partir de um conjunto reduzido de parâmetros especificados pelo professor, além de

que o modelo já tinha sido avaliado com alunos simulados, e mostrou um bom desempenho de diagnóstico tanto em termos de precisão como de eficiência e também com alunos reais [4].

1.3 Objetivos gerais do trabalho

O principal objetivo deste trabalho é o de descrever a aplicabilidade de um modelo de aluno já implementado na universidade de Aveiro e que se baseia em redes bayesianas. Descreveremos, assim, a aplicação *web* SIACUA e conceitos relacionados com este modelo, bem como a estrutura do modelo do aluno, a construção e definição de redes bayesianas, que por conseguinte requer conhecimentos de grafos para a sua representação, e de teoria de probabilidades para a sua definição, modelação e implementação.

Dedicamos grande parte do trabalho ao estudo de redes bayesianas, uma vez que estas são a base onde assenta esta aplicação. Permitem lidar com incertezas, utilizando probabilidades. A sua representação gráfica explicita as relações entre as variáveis em análise e constitui uma poderosa ferramenta na representação de conhecimentos. O nome bayesiana decorre do facto de que a distribuição de probabilidades conjunta representada por uma rede bayesiana é subjetiva e esta distribuição de probabilidades subjetiva pode ser atualizada à luz de novas evidências usando o teorema de Bayes. Assim, a probabilidade de um evento é uma medida subjetiva pessoal de crença em que o evento ocorra. Estas medidas são estimadas tendo em conta as frequências relativas de dados anteriores e/ou a opinião empírica dos professores [7].

Uma rede bayesiana, segundo Castillo et al. [3], é um formalismo que mistura a teoria dos grafos e a teoria da probabilidade. Nesse sentido, uma RB tem dois componentes principais: uma estrutura que define relacionamento qualitativo causal entre as variáveis, sendo a estrutura um grafo, e parâmetros numéricos que quantificam a relação probabilística causal entre as variáveis, que são representadas pelos nós da estrutura.

1.4 Organização da dissertação

Esta dissertação está dividida em sete capítulos, sendo estes relativos aos objetivos referidos anteriormente. Assim sendo, no segundo capítulo, apresentamos alguns conceitos da teoria de grafos que serve de base para a compreensão da estrutura das redes bayesianas. No terceiro capítulo, introduzimos alguns conceitos da teoria de probabilidades considerados relevantes para o entendimento desta dissertação, mais especificamente, o cálculo de probabilidades condicionadas e o teorema de Bayes, que será fulcral para o entendimento da propagação das probabilidades na aplicação estudada. Seguidamente, no quarto capítulo, descrevemos, de uma forma superficial, em que consiste uma rede bayesiana, que vagamente será a interação entre teoria das probabilidades e teoria dos grafos. Primeiramente, apresentamos redes bayesianas de forma intuitiva e posteriormente uma definição mais formal. Abordamos no quinto capítulo a aplicação do estudo de redes bayesianas ao problema do modelo do aluno, a propagação de probabilidades associadas aos nós das questões e a forma de simplificar a especificação dos parâmetros. As probabilidades condicionadas necessárias a definir *à priori* para cada nó, crescem exponencialmente com o números de nós na rede, mas, usando o modelo, podem ser obtidas a partir de um conjunto de parâmetros fáceis de especificar. No sexto capítulo, apresentamos a estrutura e as funcionalidades da aplicação SIACUA. Descrevemos o modelo estrutural usado para realizar o modelo de aluno baseado em redes bayesianas, assim como, descreveremos os tipos de nós definidos, as relações entre eles, neste modelo, e os parâmetros a especificar. Através de vários exemplos, simulamos a propagação das probabilidades, dadas as evidências pelos alunos, ou seja, como as probabilidades são atualizadas na rede depois do aluno responder a cada questão. Exploramos, também, algumas perspectivas quanto a uma futura melhoria da aplicação com a introdução de novos objetivos e avaliamos a eficácia da aplicação, demonstrando o funcionamento do modelo implementado. Finalmente, o capítulo sete é dedicado às principais conclusões obtidas com a elaboração deste trabalho.

Capítulo 2

Grafos

Os grafos são de grande utilidade, uma vez que são usados para modelar problemas do dia-a-dia, nas mais diversas áreas, nomeadamente na matemática, informática, engenharias e em muitas aplicações industriais, tendo assim importantes aplicações práticas. Intuitivamente, um grafo é uma representação gráfica de algum problema ou processo em que cada ponto representa uma das etapas desse problema ou processo e existem ligações entre esses pontos, permitindo, assim, uma melhor visualização de todas as possibilidades. Com esta visão geral (organização) do problema, é mais fácil obter informações sobre ele.

Neste capítulo, baseando-nos em [1] e [17], fazemos uma breve apresentação de noções básicas de grafos e das suas principais características, através da definição de alguns conceitos, como caminho, circuito, ciclo, árvore, que serão importantes para o estudo de redes bayesianas num modelo de aluno, o objetivo central deste trabalho.

Definição 2.1. Um *grafo* é um par de conjuntos (V, E) , tal que $V = V(G) = \{v_1, \dots, v_n\}$ é o conjunto dos *vértices* e $E = E(G)$ é o conjunto das *arestas*, a cada uma das quais corresponde um subconjunto de $V(G)$ de cardinalidade 2, i.e., $E(G) = \{e_1, \dots, e_m\}$, com $e_k = \{v_{k_i}, v_{k_j}\}$, para cada $k \in \{1, \dots, m\}$.

Para simplicidade de notação, uma aresta entre os vértices x e y será representada por xy .

Os grafos são muitas vezes representados por figuras planas constituídas por linhas e pontos, as primeiras representando arestas e os segundos vértices.

Exemplo 2.1.

Na figura seguinte está representado um grafo $G(V, E)$ com $V = \{A, B, C, D, F\}$ é o conjunto de vértices e $E = \{e_1, e_2, e_3, e_4, e_5\}$ é o conjunto de arestas.

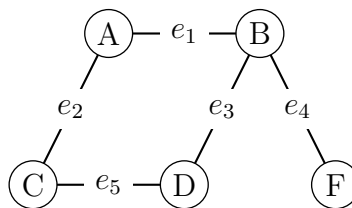


Figura 2.1: Grafo $G(V, E)$, com $V = \{A, B, C, D, F\}$ e $E = \{e_1, e_2, e_3, e_4, e_5\}$.

Um *grafo* diz-se *simples* se não existem arestas paralelas (mais do que uma aresta entre os mesmos dois vértices) nem lacetes (arestas com ambos os extremos no mesmo vértice).

Os grafos simples são designados apenas por grafos e os grafos com lacetes e/ou arestas paralelas por *multigrafos*.

Exemplo 2.2.

O grafo apresentado no exemplo 2.1 é um grafo simples, enquanto o grafo apresentado na seguinte figura (figura 2.2) é um multigrafo.

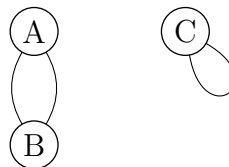


Figura 2.2: Multigrafo. AB e BA são arestas paralelas e CC é um lacete.

O número de vértices de um grafo designa-se por *ordem* do grafo. O número de arestas de um grafo designa-se por *dimensão* do grafo.

Exemplo 2.3.

O grafo do exemplo 2.1 tem ordem cinco ($|V| = 5$) e dimensão cinco ($|E| = 5$).

Definição 2.2. Dado um grafo (ou multigrafo) G designa-se por *orientação de G* e denota-se por \vec{G} a substituição das arestas de G por arcos. Assim, para todo $x, y \in V(G)$ tem-se $(x, y) \in A(\vec{G})$ ou $(y, x) \in A(\vec{G})$. O conjunto dos arcos obtido denota-se, usualmente, por $A(\vec{G})$.

Definição 2.3. Um grafo (ou digrafo) orientado também se designa por *digrafo* e também se denota por DG .

Exemplo 2.4.

Voltando ao grafo apresentado no exemplo 2.1, se substituirmos as arestas de G por arcos obtemos o seguinte digrafo.

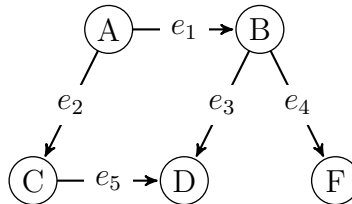


Figura 2.3: Digrafo $G(V, \vec{A})$, com $V = \{A, B, C, D, F\}$ e $\vec{A} = \{e_1, e_2, e_3, e_4, e_5\}$.

Dado um grafo G , uma aresta diz-se *incidente* no vértice v , se v é um dos seus extremos, e dois vértices, x e y , dizem-se *adjacentes* se $xy \in E(G)$.

De modo análogo, dado um digrafo \vec{G} , dois vértices, x e y , dizem-se *adjacentes* se $\{(x, y), (y, x)\} \cap A(\vec{G}) \neq \emptyset$.

Exemplo 2.5.

No grafo do exemplo 2.1 a aresta e_1 é incidente no vértice A , C e D são vértices adjacentes. No digrafo do exemplo 2.4 A e B são vértices adjacentes.

Definição 2.4. O conjunto de vértices adjacentes a um vértice v designa-se por *vizinhança de v* (ou *conjunto de vizinhos de v*) e denota-se por $N_G(v)$.

Exemplo 2.6.

No grafo do exemplo 2.1 $N_G(D) = \{B, C\}$ e $N_G(F) = \{B\}$.

Definição 2.5. Dados dois grafos G e G' , diz-se que G' é um *subgrafo* de G e que G é um *supergrafo* de G' quando $V(G') \subseteq V(G)$ e $E(G') \subseteq E(G)$.

Exemplo 2.7.

Na figura seguinte apresentamos o grafo G e o grafo H . O grafo G é um supergrafo de H e o grafo H é um subgrafo de G .

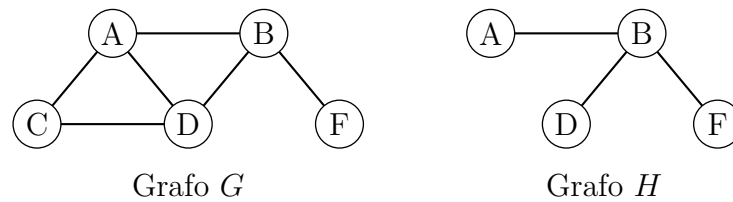


Figura 2.4: Supergrafo e subgrafo.

Definição 2.6. Dado um grafo G , designa-se por *subgrafo* de G *induzido* pelo subconjunto de vértices V' , e denota-se por $G[V']$, o subgrafo obtido de G ignorando o subconjunto de vértices $V(G) \setminus V'$ e, conseqüentemente, as arestas que lhe são incidentes.

Exemplo 2.8.

Consideremos o grafo G do exemplo anterior (exemplo 2.7) e o subconjunto de vértices $V' = \{A, B, D, F\}$, apresentamos na figura seguinte o grafo $G[V']$, subgrafo de G induzido por V' .

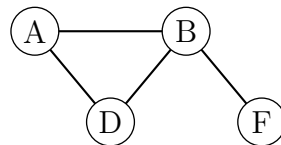
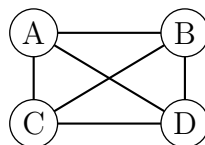


Figura 2.5: Subgrafo induzido.

Designa-se por *grafo completo* de ordem n e denota-se por K_n um grafo com n vértices dois a dois adjacentes.

Exemplo 2.9.

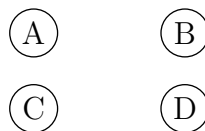
Na figura 2.6 apresentamos o grafo K_4 .

Figura 2.6: Grafo K_4 .

Designa-se por *grafo nulo* de ordem n e denota-se por N_n um grafo com n vértices dois a dois não adjacentes, ou seja, sem qualquer aresta.

Exemplo 2.10.

Na figura 2.7 apresentamos o grafo N_4 .

Figura 2.7: Grafo N_4 .

Definição 2.7. Dado um grafo G e um vértice $v \in V(G)$ designa-se por *grau* ou *valência* de v , e denota-se por $d_G(v)$, o número de arestas de G incidentes em v .

Definição 2.8. O *máximo grau* dos vértices de G denota-se por $\Delta(G)$ e é o grau dos vértice de maior grau.

Definição 2.9. O *mínimo grau* dos vértices de G denota-se por $\delta(G)$ e é o grau dos vértice de menor grau.

Exemplo 2.11.

Voltando ao grafo do exemplo 2.1 temos:

- $d_G(F) = 1$;
- $d_G(A) = d_G(C) = d_G(D) = 2$;
- $d_G(B) = 3$.

Consequentemente, $\Delta(G) = 3$ e $\delta(G) = 1$.

Definição 2.10. Dados dois grafos G e H , a *união de G com H* denota-se por $G \cup H$ e corresponde ao grafo $G \cup H = (V(G) \cup V(H), E(G) \cup E(H))$ e a *interseção de G com H* denota-se por $G \cap H$ e corresponde ao grafo $G \cap H = (V(G) \cap V(H), E(G) \cap E(H))$.

Exemplo 2.12.

Na figura seguinte apresentamos os grafos G , H , $G \cup H$ e $G \cap H$.

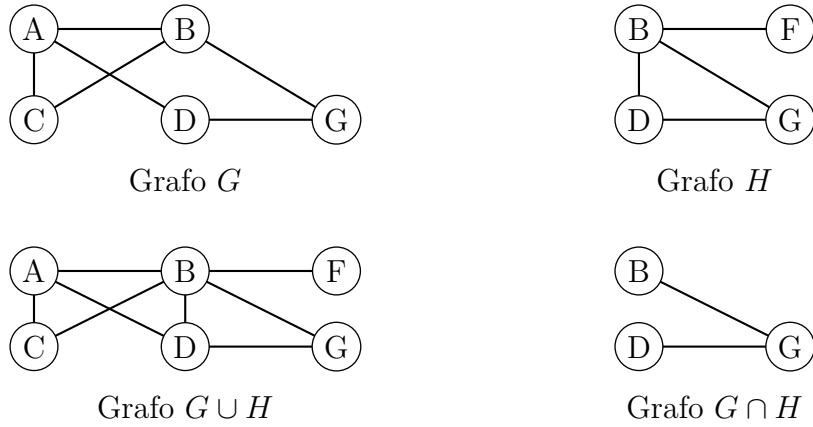


Figura 2.8: União e interseção de grafos.

Definição 2.11. Designa-se por *passeio* num grafo, entre os vértices x e y , toda a sequência de vértices e arestas da forma $x = v_1, v_1v_2, v_2, \dots, v_{k-1}, v_{k-1}v_k, v_k = y$ com eventual repetição de vértices e arestas. Neste caso, os vértices x e y designam-se por *vértices extremos* do passeio (sendo x o *vértice inicial* e y o *vértice final*).

Um *trajeto* num grafo G entre os vértices x e y é um passeio entre x e y sem arestas repetidas (podendo, no entanto, existir vértices repetidos).

Um *caminho* entre os vértices x e y é um trajeto entre x e y sem vértices repetidos.

Os trajetos fechados (onde o vértice final coincide com o inicial) designam-se por *circuitos*.

Os trajetos fechados, onde os vértices inicial e final são os únicos que coincidem, designam-se por *ciclos*.

Geralmente, os caminhos, trajetos, passeios, circuitos e ciclos representam-se pela respetiva sequência de vértices.

Exemplo 2.13.

Consideremos o grafo representado na figura 2.9.

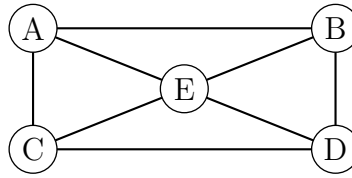


Figura 2.9: Grafo G do exemplo 2.13.

- Um passeio entre o vértice A e o vértice D pode ser representado por A, AB, BE, EA, AB, BD, D ou por A, B, E, A, B, D ;
- A, B, E, C, A, E, D é um trajeto entre os vértices A e D ;
- A, B, D é um caminho entre os vértices A e D ;
- A, B, E, D, C, E, A é um circuito;
- A, B, D, C, A é um ciclo.

Definição 2.12. Dado um caminho P de um grafo G designa-se por comprimento de P e denota-se por $comp(P)$ o número de arestas que o constitui ($comp(P) = |E(P)|$).

Definição 2.13. Dado um grafo G de ordem n e dois vértices $x, y \in V$, denotando por $\mathcal{P}_G(x, y)$ o conjunto de todos os caminhos de G entre x e y , designa-se por *distância* entre vértices de G a função

$$d_G : V \times V \longrightarrow \{0, \dots, n-1\} \cup \{\infty\}$$

$$(x, y) \longrightarrow d_G(x, y) = \begin{cases} \min\{comp(P) : P \in \mathcal{P}_G(x, y)\} & \text{se } \mathcal{P}_G(x, y) \neq \emptyset \\ \infty & \text{se } \mathcal{P}_G(x, y) = \emptyset \end{cases}$$

Exemplo 2.14.

Consideremos o grafo representado na figura 2.10 e seja $P = A, B, D, C$ um caminho entre os vértices A e C .

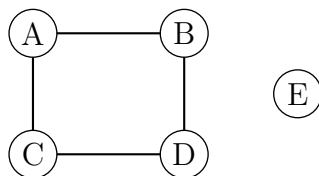


Figura 2.10: Grafo G do exemplo 2.14.

Temos que $\text{comp}(P) = 3$, $d_G(A, C) = 1$ e $d_G(B, E) = \infty$.

Definição 2.14. Um grafo G diz-se *conexo* se existir um caminho entre quaisquer dois dos seus vértices.

Exemplo 2.15.

O grafo apresentado no exemplo 2.13 é um grafo conexo e o grafo apresentado no exemplo 2.14 não é conexo.

Definição 2.15. Os grafos conexos acíclicos (i.e, sem ciclos) designam-se por *árvores*.

Exemplo 2.16.

O grafo apresentado na figura 2.11 é uma árvore.

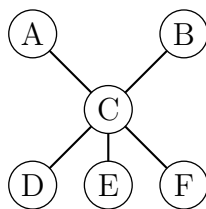


Figura 2.11: Uma árvore.

Definição 2.16. Designa-se por *floresta* todo o grafo acíclico.

Exemplo 2.17.

O grafo apresentado na figura 2.12 é uma floresta.

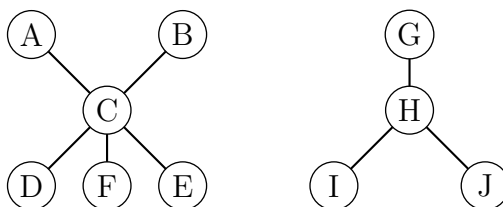


Figura 2.12: Uma floresta.

Teorema 2.1. *Seendo G um grafo, são equivalentes as seguintes afirmações:*

1. G é uma árvore;
2. G é conexo e tem $|V(G) - 1|$ arestas;
3. G não tem circuitos, mas acrescentando-se uma aresta a G resulta um único circuito.

Demonstração.

Implicação $1 \Rightarrow 2$

Dado que uma árvore é um grafo conexo, basta provar que se $G = (V(G), E(G))$ é uma árvore, então tem $|V(G)| - 1$ arestas. Vamos fazer esta prova por indução sobre o número de arestas de grafos que definem árvores.

Seja $G_k = (V(G_k), E(G_k))$ uma árvore com k arestas.

Para $k = 1$ vem $|V(G_k)| = 2$, pelo que o resultado se verifica.

Suponha-se que o resultado é verdadeiro para k tal que $1 \leq k \leq n - 1$ e considere-se a árvore G_n . Dado que G_n não tem circuitos, existe pelo menos um vértice v , com grau 1. Considerando o subgrafo obtido de G_n , retirando-se o vértice v (e consequentemente a aresta que lhe é incidente), determina-se G_{n-1} que continua a ser uma árvore (uma vez que permanece conexo e sem circuitos), logo, por hipótese de indução $|E(G_{n-1})| = |V(G_{n-1})| - 1$. Dado que $|V(G_n)| = |V(G_{n-1})| + 1$ conclui-se que G_n tem $|V(G_n)| - 1$ arestas, completando-se assim a prova da implicação $1 \Rightarrow 2$.

Implicação $2 \Rightarrow 3$

Suponhamos que G é conexo e tem $|V(G)| - 1$ arestas. Como G é conexo todos os vértices tem pelo menos grau 1. Uma vez que G tem $|V(G)| - 1$ arestas, é claro, que para quaisquer dois vértices diferentes de G , v_1, v_2 temos $N_G(v_1) \cap N_G(v_2) = \emptyset$. Logo G não tem circuitos. Como cada vértice tem pelo menos grau 1, se acrescentarmos uma aresta esta vai ligar dois vértices de G , v_1 e v_2 por exemplo. Assim $N_G(v_1) \cap N_G(v_2) = 1 \neq \emptyset$ e existe um único circuito.

Implicação $3 \Rightarrow 1$

Suponhamos que G não tem circuitos, mas que acrescentando-se uma aresta a G resulta um único circuito. É óbvio que G é acíclico pois se não tem circuitos, não tem ciclos. Resta mostrar que G é conexo.

Suponhamos que G não é conexo, ou seja, suponhamos que não existe um caminho entre dois vértices, v_1 e v_2 , de G . Ao acrescentar a aresta v_1v_2 , G continua a não ter circuitos, o que contradiz a hipótese. Logo, G é conexo e consequentemente G é uma árvore. \square

Definição 2.17. Dado um grafo conexo G , designa-se por *árvore abrangente* ou *árvore de suporte* de G todo o subgrafo de G que é uma árvore e contém todos os vértices de G .

Exemplo 2.18.

Na figura 2.13 apresentamos um grafo G e uma sua árvore abrangente.

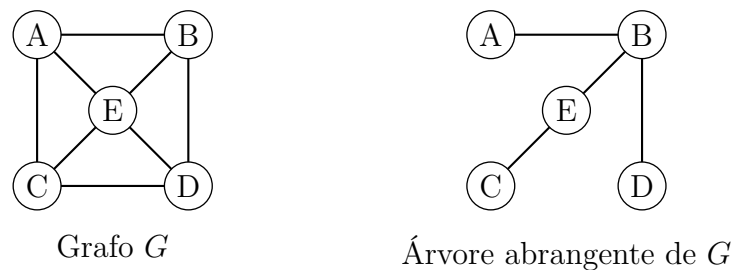


Figura 2.13: Árvore abrangente do Grafo G .

Teorema 2.2. *Todo o grafo conexo admite uma árvore abrangente.*

Demonstração.

Seja G um grafo conexo.

Se G não tem circuitos então, por definição, é uma árvore e o resultado verifica-se.

Suponha-se que G tem um circuito. Então retirando uma aresta a esse circuito o grafo mantém-se conexo.

Se G tem mais de um circuito, então ao repetirmos o processo anterior, ao fim de um número finito de arestas eliminadas, obtém-se uma árvore abrangente (uma vez que o conjunto de vértices não foi alterado). \square

Definição 2.18. Dado um digrafo G diz-se que o vértice x é um *pai* do vértice y se existe um arco de x para y . Diz-se, também, que y é filho de x e o conjunto dos pais de y representa-se por $pa(y)$.

Exemplo 2.19.

Consideremos o digrafo apresentado na figura seguinte (figura 2.14).

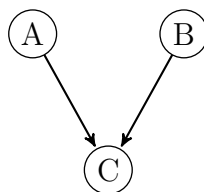


Figura 2.14: Digrafo do exemplo 2.19.

O vértice A é um pai do vértice C , C é um filho do vértice B e $pa(C) = \{A, B\}$.

Definição 2.19. Dado um digrafo G diz-se que o vértice x é um antepassado do vértice z se existir um caminho orientado de x para z . Diz-se, também, que z é o descendente de x e o conjunto dos descendentes de x denota-se por $de(x)$.

Exemplo 2.20.

Consideremos o digrafo apresentado na figura seguinte (figura 2.15).

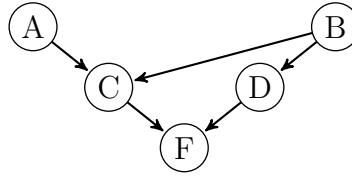


Figura 2.15: Digrafo do exemplo 2.19.

O vértice A é um antepassado do vértice F , o vértice F é um descendente do vértice B e $de(B) = \{C, D, F\}$.

Depois desta breve revisão de alguns conceitos da teoria dos grafos necessários para a compreensão da representação das redes bayesianas, estudamos, no capítulo seguinte, alguns conceitos da teoria das probabilidades necessárias para o mesmo efeito.

Capítulo 3

Probabilidades

As probabilidades são largamente utilizadas, nas mais diversas áreas, pois fornecem bases para a tomada de decisão, quando existe incerteza. Embora não consigamos prever o futuro, podemos conjecturar a probabilidade de acontecimentos, usando para esse efeito o conhecimento do passado. Na medicina, por exemplo, os médicos baseiam-se na sua experiência e nas suas convicções para quantificar a probabilidade de um doente sofrer de determinada doença dado os seus sintomas.

Um dos conceitos fundamentais da teoria das probabilidades é a regra de Bayes. Esta regra foi introduzida, por Bayes, em 1763 e permite derivar probabilidades a partir de um conjunto de outras probabilidades.

Baseando-se na documentação de [8], [11] e [21], este capítulo introduz conceitos da teoria de probabilidades, suas definições, axiomas, distribuições de probabilidade e Regra de Bayes, que constituem o alicerce fundamental para o entendimento das redes bayesianas, dado que a sua metodologia é baseada em probabilidades, especialmente a probabilidade condicionada.

3.1 Conceitos Iniciais

As experiências podem ser classificadas de duas formas: aleatórias ou deterministas.

Definição 3.1. Uma experiência, E , diz-se *aleatória* se for um procedimento que se pode repetir um grande número de vezes nas mesmas condições ou pelo menos em condições semelhantes e em que todos os resultados possíveis são conhecidos à priori, mas o resultado exato não é conhecido antes da realização da experiência, i.e., é imprevisível.

Exemplo 3.1.

Alguns exemplos de experiências aleatórias:

- “lançamento de um dado e registo do número de pontos que sai”, nesta experiência o resultado é imprevisível, pois não sabemos à priori, das seis faces, qual a que fica voltada para cima;
- “lançamento de duas moedas”, nesta experiência, também, não sabemos se é a face comum ou a face nacional que fica voltada para cima em cada uma das moedas;
- “lançamento de um dado e de uma moeda”;
- “lançamento de uma moeda até aparecer face”.

Definição 3.2. Uma experiência é *determinista* sempre que é possível prever o resultado que se obtém quando a experiência é repetida nas mesmas condições.

Exemplo 3.2.

Alguns exemplos de experiências deterministas:

- “colocar dinheiro num banco e calcular o juro produzido num certo tempo”, à partida já sabemos quanto iremos receber de juros;
- “colocar uma massa numa mola e medir o alongamento da mola”;
- “colocar dois produtos químicos em contacto e observar a reação”.

Definição 3.3. Designa-se por *Espaço Amostral* e representa-se por Ω , o conjunto de todos os resultados possíveis associados a uma experiência aleatória.

Exemplo 3.3.

Considerando as experiências aleatórias dos exemplos 3.1, os respetivos espaços amostrais são:

- $\Omega_1 = \{1, 2, 3, 4, 5, 6\};$
- $\Omega_2 = \{CC, FC, FF, CF\} = \{(C, C), (F, C), (F, F), (C, F)\};$
- $\Omega_3 = \{1C, 1F, 2C, 2F, \dots, 6C, 6F\};$
- $\Omega_4 = \{F, CF, CCF, CCCF, \dots\}.$

Definição 3.4. Um *acontecimento* é um qualquer subconjunto do conjunto de resultados, Ω , numa experiência aleatória.

Exemplo 3.4.

No acontecimentos, *saída de face par*, numa experiência aleatória de “lançamento de um dado”, pode representar-se por $A = \{2, 4, 6\} \subset \Omega = \{1, 2, 3, 4, 5, 6\}.$

Um acontecimentos é, portanto, o conjunto de resultados de uma experiência; por vezes identificamos um acontecimento com a sua realização, ou seja, o conjunto dos seus resultados.

A importância do espaço amostra Ω advém especialmente da sua utilidade na definição de acontecimentos. Normalmente há maior interesse nos acontecimentos e nas famílias de acontecimentos do que propriamente nos elementos do espaço Ω .

Um acontecimento designa-se *simples* ou *elementar* quando contém apenas um resultado, ou seja, é um conjunto singular e *composto* quando contém mais do que um resultado. Ω é o acontecimento *certo*, pois é aquele que ocorre sempre. Um acontecimento *impossível* é aquele que nunca ocorre, ou seja, é o conjunto vazio e representa-se através do símbolo \emptyset . Dado que os acontecimentos são conjuntos (em particular, subconjuntos do espaço

amostral), podem aplicar-se-lhes as operações de álgebra de conjuntos, em particular, de reunião, interseção e complementaridade.

Exemplo 3.5.

Na experiência aleatória “tirar uma bola de um saco com 10 bolas numeradas de 1 a 10”, o espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

- O acontecimento *sair uma bola com um número natural menor que onze* é Ω ;
- O acontecimentos *sair uma bola com um número negativo* é impossível;
- O acontecimento *sair a bola com o número 7* (que se representa por $\{7\}$) é elementar;
- *Sair uma bola com um número par* é um acontecimento composto, $\{2, 4, 6, 8, 10\}$.

Para facilitar a exposição supõe-se Ω equivalente a um retângulo do espaço \mathbb{R}^2 e os acontecimentos seus subconjuntos, representados por diagramas de Venn. Apresentamos de seguida alguns resultados (propriedades e definições) relativos à álgebra dos acontecimentos.

Definição 3.5. Dois acontecimentos, A e B, dizem-se *mutuamente exclusivos* se não puderem ocorrer simultaneamente; em linguagem de conjuntos escreve-se $A \cap B = \emptyset$. Mais geralmente, $n(n > 2)$ acontecimentos A_1, A_2, \dots, A_n dizem-se mutuamente exclusivos se $A_i \cap A_j = \emptyset$ sempre que $i \neq j$ e $1 \leq i, j \leq n$.

Exemplo 3.6.

Na experiência aleatória “tirar uma bola de um saco com 10 bolas numeradas de 1 a 10”, os acontecimentos A: *sair um número menor do que três* e B: *sair o número quatro*, são acontecimentos mutuamente exclusivos, $A = \{1, 2\}$ e $B = \{4\}$ e $A \cap B = \emptyset$.

Definição 3.6. Dois acontecimentos, A e B, dizem-se contrários se $A \cap B = \emptyset$ e $A \cup B = \Omega$, e representa-se o acontecimento contrário de A por \bar{A} .

Exemplo 3.7.

Na experiência aleatória “tirar uma bola de um saco com 10 bolas numeradas de 1 a 10”, os acontecimentos A: *sair um número par* e B: *sair um número ímpar*, são acontecimentos contrários, uma vez que $A = \{2, 4, 6, 8, 10\}$, $B = \{1, 3, 5, 7, 9\}$, tendo-se $A \cap B = \emptyset$ e $A \cup B = \Omega$.

Na tabela 3.1 apresentam-se algumas propriedades da interseção e da união de acontecimentos.

PROPRIEDADES	UNIÃO	INTERSEÇÃO
1. Comutativa	$A \cup B = B \cup A$	$A \cap B = B \cap A$
2. Associativa	$A \cup (B \cup C) =$ $= (A \cup B) \cup C$	$A \cap (B \cap C) =$ $= (A \cap B) \cap C$
3. Distributiva	$A \cup (B \cap C) =$ $= (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) =$ $= (A \cap B) \cup (A \cap C)$
4. Idempotência	$A \cup A = A$	$A \cap A = A$
5. Lei do complemento	$A \cup \bar{A} = \Omega$	$A \cap \bar{A} = \emptyset$
6. Leis de De Morgan	$\overline{A \cup B} = \bar{A} \cap \bar{B}$	$\overline{A \cap B} = \bar{A} \cup \bar{B}$
7. Elemento neutro	$A \cup \emptyset = A$	$A \cap \Omega = A$
8. Elemento absorvente	$A \cup \Omega = \Omega$	$A \cap \emptyset = \emptyset$

Tabela 3.1: Propriedades das operações entre acontecimentos.

Definiu-se acontecimento de uma forma genérica como um subconjunto do espaço amostra Ω associado a uma experiência aleatória. No caso do número de resultados do espaço ser finito, podem-se definir acontecimentos em número finito, em particular, se Ω contiver n elementos (ou seja, n resultados elementares) então existem 2^n acontecimentos.

Quando o número de resultados em Ω não for finito é necessário trabalhar com uma coleção de subconjuntos de Ω , a qual designaremos por F , mais restrita do que a classe de todos os subconjuntos de Ω . A essa coleção de subconjuntos probabilizáveis chamamos

acontecimentos. Essa classe F deve ter as propriedades essenciais da classe de todos os subconjuntos de Ω , sendo elas:

- I. O espaço $\Omega \in F$;
- II. O acontecimento impossível $\emptyset \in F$;
- III. Se A_1, A_2, \dots , forem acontecimentos de F em número finito ou infinito numerável então $\bigcup A_i \in F$;
- IV. $A, B \in F \Rightarrow A \setminus B \in F$;
- V. Se A_1, A_2, \dots , forem acontecimentos de F em número finito ou infinito numerável então $\bigcap A_i \in F$.

Definição 3.7. F é uma σ -Álgebra se e só se F é uma classe não vazia tal que:

- I. Se A_1, A_2, \dots , forem acontecimentos de F em número finito ou infinito numerável, então $\bigcup A_i \in F$;
- II. $A \in F \Rightarrow \bar{A} \in F$.

Facilmente se verifica que as propriedades indicadas da classe de todos os subconjuntos de Ω e que deverão ser mantidas para a classe F , são asseguradas se esta for uma σ -Álgebra.

Exemplo 3.8.

Exemplos de σ -Álgebra:

1. Conjunto das partições de Ω é a maior σ -Álgebra que contém Ω . Se $\#\Omega = n$, então existem 2^n subconjuntos de Ω .
2. $\{\emptyset, \Omega\}$ é a menor σ -Álgebra que contém Ω .
3. $\{\emptyset, \Omega, A, \bar{A}\}$ é a menor σ -Álgebra que contém o conjunto A .
4. $\Omega = \mathbb{R}$. Uma definição corrente de acontecimentos na reta são intervalos. Considera-se, então, a σ -Álgebra gerada por intervalos do tipo: $I =]a, b] = \{x : a < x \leq b\}$.

Definição 3.8. Chama-se *corpo de Borel* em \mathbb{R} e por \mathcal{B} , à menor σ -Álgebra que contém intervalos do tipo $I =]a, b]$.

Como o corpo de Borel é um caso particular de uma σ -Álgebra, as propriedades indicadas, da classe de todos os subconjuntos de Ω , são verificadas. Pois, \mathcal{B} contém todas as uniões numeráveis de I , contém todas as interseções numeráveis de I e contém todos os complementares dos intervalos do tipo I . Contém, ainda, conjuntos reais singulares (só com um único elemento), pois se considerarmos a interseção de um número infinito numerável de intervalos do tipo $I_n = \{x : a - \frac{1}{n} < x \leq a\}$ temos que $\bigcap_{n=1}^{\infty} I_n = \{a\}$. Consequentemente também pertencem a \mathcal{B} intervalos fechados, pois $[a, b] = \{a\} \cup]a, b]$ e intervalos abertos, uma vez que $]a, b[=]a, b] \setminus \{b\}$.

3.2 Cálculo de Probabilidades. Lei de Laplace

A cada experiência aleatória podemos associar uma infinidade de acontecimentos. Para distinguirmos os vários acontecimentos, torna-se necessário associar a cada acontecimento A , um número que de alguma maneira medirá o quanto verosímil é que o acontecimento A venha a ocorrer. Este número é a probabilidade do acontecimento A , $P(A)$.

Começemos com uma noção intuitiva do conceito de probabilidade de um acontecimento.

Consideremos uma urna que contem 49 bolas azuis e 1 bola branca. Se efetuarmos uma extração, teremos duas possibilidades: bola azul ou bola branca. Percebemos, entretanto, que será muito mais frequente obtermos numa extração uma bola azul. Podemos afirmar que o acontecimento *sair bola azul* é mais provável, ou tem maior probabilidade de ocorrer do que o acontecimento *sair bola branca*.

A primeira definição de probabilidade conhecida deve-se a Laplace no início do séc. XIX, para a hipótese de casos igualmente prováveis ou o chamado princípio da simetria.

Definição 3.9. Lei de Laplace

A probabilidade de um acontecimento A é o quociente entre o número de casos favoráveis à

realização desse acontecimento (N_A) e o número total de casos possíveis ($N = |\Omega|$), desde que todos os casos sejam igualmente prováveis.

$$P(A) = \frac{N_A}{N}.$$

O que a definição de Laplace nos diz é que se todos os resultados, forem igualmente verosímeis, a probabilidade de cada resultado, i.e., a probabilidade de cada acontecimento elementar A_i , $1 \leq i \leq N$ é:

$$P(A_i) = \frac{1}{N}, \forall i = 1, \dots, N.$$

Exemplo 3.9.

A probabilidade de *sair face par* no caso do “lançamento de um dado” é $\frac{3}{6} = \frac{1}{2}$.

Como resultado da definição acima apresentada, as probabilidades satisfazem alguns axiomas referidos de seguida.

Definição 3.10. Definição axiomática de probabilidade

Considere-se uma experiência aleatória cujo espaço dos resultados é Ω (finito). Chama-se probabilidade a uma aplicação que a cada acontecimento de Ω faz corresponder um número real que satisfaz os seguintes axiomas:

- **Axioma 1:** $P(A) \geq 0$, para todo $A \subset \Omega$;
- **Axioma 2:** $P(\Omega) = 1$;
- **Axioma 3:** $P(A \cup B) = P(A) + P(B)$, se $A \cap B = \emptyset$.

Generalizando o axioma 3, se A_1, A_2, \dots, A_n , são acontecimentos mutuamente exclusivos definidos em Ω , tem-se que:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

Com base nos axiomas anteriores, podem deduzir-se propriedades associadas às probabilidades interessantes. Apresentam-se seguidamente três exemplos:

Teorema 3.1. Para qualquer acontecimento A com probabilidade $P(A)$,

$$P(A) + P(\bar{A}) = 1.$$

Demonstração.

Como A e \bar{A} são acontecimentos mutuamente exclusivos, aplicando os axiomas 3 e 2, sucessivamente, resulta $P(A) + P(\bar{A}) = P(A \cup \bar{A}) = P(\Omega) = 1$. \square

Teorema 3.2. $P(\emptyset) = 0$.

Demonstração.

Como \emptyset e Ω são acontecimentos complementares, o resultado obtém-se imediatamente do teorema 3.2 e do axioma 2. \square

Teorema 3.3. Para quaisquer acontecimentos A e B ,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Demonstração.

$A \cup B = A \cup (\bar{A} \cap B)$, pois $A \cup (\bar{A} \cap B) = (A \cup \bar{A}) \cap (A \cup B) = \Omega \cap (A \cup B) = A \cup B$ e os acontecimentos A e $\bar{A} \cap B$ são mutuamente exclusivos.

Do axioma 3 resulta que $P(A \cup B) = P(A) + P(\bar{A} \cap B)$.

Ora o acontecimento B pode ser expresso como reunião de dois acontecimentos mutuamente exclusivos, designadamente $B = (A \cap B) \cup (\bar{A} \cap B)$.

Daqui vem que $P(B) = P(A \cap B) + P(\bar{A} \cap B) \Leftrightarrow P(\bar{A} \cap B) = P(B) - P(A \cap B)$.

Substituindo esta expressão na anteriormente obtida para $P(A \cup B)$, obtém-se

$$P(A \cup B) = P(A) + P(\bar{A} \cap B) = P(A) + P(B) - P(A \cap B).$$

\square

3.3 Probabilidade Condicionada

O conceito de probabilidade condicionada é um dos mais importantes da teoria das probabilidades, pois em muitas situações em que se pretende calcular a probabilidade de

um acontecimento, já se dispõe de alguma informação sobre o resultado da experiência, a qual permite atualizar a atribuição de probabilidades a esse acontecimento.

Consideremos o exemplo 3.5, definido anteriormente, que consiste na em tirar uma bola de um saco com 10 bolas numeradas de 1 a 10. A probabilidade do acontecimentos A: *sair número 2*, é $\frac{1}{10}$. Se, no entanto, pretendemos a probabilidade desse mesmo acontecimento sabendo de antemão que o acontecimento B: *sair um número par* se realizou, o espaço de resultados é constituído por cinco resultados igualmente possíveis, dos quais um é favorável, pelo que a probabilidade pretendida é $\frac{1}{5}$, o dobro da obtida anteriormente, quando não tínhamos nenhuma informação. A esta última probabilidade chamamos probabilidade condicionada A sabendo que se realizou o acontecimento B e representa-se por $P(A|B)$.

Definição 3.11. Dados dois acontecimentos A e B, com $P(B) > 0$, define-se *probabilidade de A sabendo que B ocorreu* e representa-se por $P(A|B)$, como sendo

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (3.1)$$

Teorema 3.4. Se $P(A) > 0$ e $P(B) > 0$ então

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A). \quad (3.2)$$

Demonstração.

Atendendo à definição de probabilidade condicionada (definição 3.11):

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A \cap B) = P(B)P(A|B). \\ P(B|A) &= \frac{P(A \cap B)}{P(A)} \Leftrightarrow P(A \cap B) = P(A)P(B|A). \end{aligned}$$

□

Generalizando a três acontecimentos, sendo $P(A) > 0$ e $P(A \cap B) > 0$:

$$\begin{aligned} P(A \cap B \cap C) &= P((A \cap B) \cap C) \\ &= P(C|(A \cap B)) \cdot P(A \cap B) \\ &= P(C|(A \cap B)) \cdot P(B|A) \cdot P(A). \end{aligned}$$

Generalizando a $n \geq 3$ acontecimentos A_1, A_2, \dots, A_n , tem-se

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_n|(A_1 \cap A_2 \cap \dots \cap A_{n-1})) \cdot \\ &\quad \cdot P(A_{n-1}|(A_1 \cap A_2 \cap \dots \cap A_{n-2})) \cdot \\ &\quad \dots P(A_2|A_1) \cdot P(A_1). \end{aligned}$$

esta fórmula designa-se por *Fórmula das probabilidades compostas*.

Exemplo 3.10.

Numa universidade os alunos que não sejam aprovados num primeiro exame a uma determinada disciplina podem realizar um segundo exame. O João tem 75% de êxito no primeiro exame e tem 90% de êxito no segundo exame. Pretendemos determinar a probabilidade do João reprovar a essa disciplina. Consideremos os acontecimentos:

A: ter êxito no primeiro exame.

B: ter êxito no segundo exame.

Sabemos que $P(A) = 0.75$ e que $P(B|\bar{A}) = 0.9$ (o João só realiza o segundo exame se reprovar no primeiro), logo

$$P(\bar{A}) = 1 - 0.75 = 0.25.$$

$$P(\bar{B}|\bar{A}) = 1 - 0.9 = 0.1.$$

Pretendemos determinar a probabilidade do acontecimento $\bar{A} \cap \bar{B}$ (o João não tem êxito no primeiro exame nem no segundo).

$$P(\bar{A} \cap \bar{B}) = P(\bar{B}|\bar{A}) \times P(\bar{A}) = 0.1 \times 0.25 = 0.025.$$

3.4 Acontecimentos Independentes

A realização de um acontecimento pode não ser afetada pela realização de outro acontecimento, ou seja, a probabilidade de um ocorrer não se altera com o facto do outro ocorrer. Dois acontecimentos nestas condições dizem-se independentes.

Definição 3.12. Dois acontecimentos A e B , $P(A) > 0$ e $P(B) > 0$, dizem-se *independente*, se a probabilidade de A se verificar, é igual à probabilidade condicionada de A se realizar, dado que B se realizou

$$P(A) = P(A|B). \quad (3.3)$$

Equivalentemente, podemos dizer que A e B , sendo $P(A) > 0$ e $P(B) > 0$, são independentes se e só se a probabilidade da sua interseção for igual ao produto das probabilidades de cada um deles, ou seja, se a seguinte condição se verificar:

$$P(A \cap B) = P(A) \cdot P(B). \quad (3.4)$$

Segundo a definição de independência, dois acontecimentos não podem ser simultaneamente independentes e mutuamente exclusivos. Se A e B são mutuamente exclusivos, então $A \cap B = \emptyset$, pelo que $P(A \cap B) = 0$, donde pela definição de independência teríamos $P(A)P(B) = 0$, que seria equivalente a ter $P(A) = 0$ ou $P(B) = 0$.

Esta condição pode ser generalizada a n elementos.

Definição 3.13. Os acontecimentos A_1, A_2, \dots, A_n , com $P(A_i) > 0$ para $i = 1, \dots, n$, dizem-se *independentes* se e só se verificarem as seguintes restrições:

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \text{ para } i \neq j$$

$$P(A_i \cap A_j \cap A_k) = P(A_i) \cdot P(A_j) \cdot P(A_k) \text{ para } i \neq j, i \neq k, j \neq k$$

...

$$P\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n P(A_k).$$

Exemplo 3.11.

Consideremos a experiência aleatória que consiste na extração de uma bola de uma urna em que estão 20 bolas, numeradas de 1 a 20. Extraí-se então uma bola da urna e verifica-se que o seu número é múltiplo de 3. Pretendemos determinar a probabilidade de a bola extraída ter um número par. O espaço amostral desta experiência é o conjunto $\Omega = \{1, 2, \dots, 20\}$.

Sejam os acontecimentos:

A: *sair bola com um número múltiplo de 3.*

B: *sair bola com um número par.*

Então, $A = \{3, 6, 9, 12, 15, 18\}$ e $B = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$. Pelo que,

$$P(A \cap B) = \{6, 12, 18\}.$$

Logo, a probabilidade de ocorrer a saída de bola com um número par, sabendo que já saiu uma bola numerada com um múltiplo de 3, é:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{3}{20}}{\frac{6}{20}} = 0.5.$$

No entanto, reparamos facilmente que $P(B) = \frac{10}{20} = 0.5$. Logo, $P(B|A) = P(B)$.

E para confirmarmos que os acontecimentos A e B são, efetivamente, independentes, basta ter em conta que $P(A \cap B) = \frac{3}{20}$, $P(A) = \frac{6}{20}$ e $P(B) = \frac{10}{20}$. Pelo que

$$P(A \cap B) = P(A) \times P(B).$$

3.5 Teorema da probabilidade total e Regra de Bayes

Em certas condições, é difícil calcular a probabilidade de um acontecimento B diretamente, mas esta poderá ser determinada, através do conhecimento de outros acontecimentos que formam uma partição do espaço amostral.

Nesta secção apresentamos o teorema da probabilidade total, a qual permite a determinação das probabilidades nestas condições. Analisamos, ainda, a Regra de Bayes, que permite inverter probabilidades condicionadas.

Definição 3.14. Os acontecimentos A_1, A_2, \dots, A_n definem uma *partição em Ω* , quando verificam, simultaneamente, as seguintes condições:

- i. A união de todos os acontecimentos é o próprio espaço amostral:

$$\bigcup_{i=1}^n A_i = \Omega;$$

- ii. Os acontecimentos são mutuamente exclusivos, dois a dois:

$$A_i \cap A_j = \emptyset \quad i \neq j, \quad i, j = 1, 2, \dots, n.$$

Na Figura 3.1 encontra-se representada uma partição de Ω .

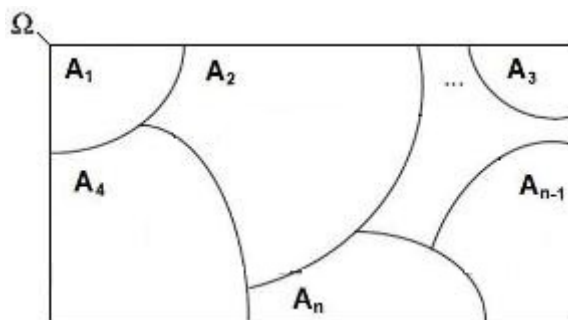


Figura 3.1: Partição de Ω .

Teorema 3.5. (Teorema da probabilidade total)

Se os acontecimentos A_1, A_2, \dots, A_n definem uma partição sobre Ω e $P(A_i) > 0$, $i = 1, \dots, n$, então para qualquer acontecimento B contido em Ω tem-se que:

$$P(B) = \sum_{i=1}^n P(B|A_i) \times P(A_i). \quad (3.5)$$

Demonstração.

Tem-se, $B = B \cap \Omega = B \cap (A_1 \cup A_2 \cup \dots \cup A_n) = B \cap (\bigcup_{i=1}^n A_i) = \bigcup_{i=1}^n (B \cap A_i)$.

Uma vez que os acontecimentos A_i são mutuamente exclusivos, os acontecimentos $(B \cap A_i)$, $i = 1, 2, \dots, n$ também o são, consequentemente

$$P(B) = P \left[\bigcup_{i=1}^n (B \cap A_i) \right] = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i) \times P(B|A_i).$$

□

Esquemáticamente,

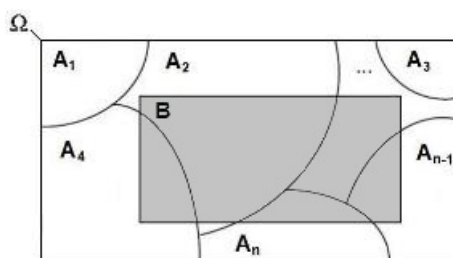


Figura 3.2: Representação do teorema da probabilidade total.

$P(B)$ vem igual à soma das probabilidades dos acontecimentos sombreados no diagrama, isto é, dos acontecimentos $(A_i \cap B)$, com $i = 1, 2, \dots, n$.

Exemplo 3.12.

Num hospital ingressaram 50% de pessoas com a doença X , 30% com a doença Y e 20% com a doença Z . A probabilidade de cura da doença X é 0,7; para as doenças Y e Z , a respetiva probabilidade é de 0,8 e 0,9. Qual é a probabilidade de um doente ter alta?

Para resolver o problema devemos em primeiro lugar definir todos os acontecimentos.

A_1 : Uma pessoa ter a doença X .

A_2 : Uma pessoa ter a doença Y .

A_3 : Uma pessoa ter a doença Z .

C : Uma pessoa ser curada.

Pretendemos calcular a $P(C)$.

Sabemos que: $P(A_1) = 0,5$; $P(A_2) = 0,3$; $P(A_3) = 0,2$; $P(C|A_1) = 0,7$; $P(C|A_2) = 0,8$ e $P(C|A_3) = 0,9$ e temos que $P(A_1) + P(A_2) + P(A_3) = 1$.

Recorrendo ao Teorema da probabilidade total:

$$\begin{aligned} P(C) &= \sum_{i=1}^3 P(C|A_i) \times P(A_i) \\ &= P(C|A_1) \times P(A_1) + P(C|A_2) \times P(A_2) + P(C|A_3) \times P(A_3) \\ &= 0,7 \times 0,5 + 0,8 \times 0,3 + 0,9 \times 0,2 \\ &= 0,77. \end{aligned}$$

Teorema 3.6. (Regra de Bayes)

Se os acontecimentos A_1, A_2, \dots, A_n definem uma partição sobre Ω e $P(A_i) > 0, i = 1, \dots, n$, então para qualquer acontecimento B , tal que $P(B) > 0$ definido em Ω tem-se que:

$$P(A_j|B) = \frac{P(A_j) \times P(B|A_j)}{\sum_{i=1}^n P(A_i) \times P(B|A_i)}, \quad \text{para } j = 1, 2, \dots, n.$$

Demonstração.

De facto,

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j) \times P(B|A_j)}{\sum_{i=1}^n P(A_i) \times P(B|A_i)}.$$

em que o numerador se obtém pela definição de probabilidade da interseção de dois acontecimentos e o denominador obtém-se pelo teorema da probabilidade total. \square

Exemplo 3.13.

Voltando ao exemplo anterior e sabendo que a um doente internado foi dado alta, calcule a probabilidade de que essa pessoa tenha sofrido da doença X . Pretendemos calcular $P(A_1|C)$. Recorrendo à Regra de Bayes:

$$\begin{aligned} P(A_1|C) &= \frac{P(A_1) \times P(C|A_1)}{\sum_{i=1}^3 P(A_i) \times P(C|A_i)} \\ &= \frac{P(A_1) \times P(C|A_1)}{P(A_1) \times P(C|A_1) + P(A_2) \times P(C|A_2) + P(A_3) \times P(C|A_3)} \\ &= \frac{0,5 \times 0,7}{0,77} \\ &= \frac{0,35}{0,77} \\ &= \frac{5}{11}. \end{aligned}$$

3.6 Distribuição de Probabilidade

Anteriormente definimos experiência aleatória, como um procedimento que se pode repetir um grande número de vezes nas mesmas condições em que todos os resultados possíveis são conhecidos à priori, mas o resultado exato não é conhecido antes da realização da experiência. Uma variável aleatória associa um número real a cada resultado dessa experiência aleatória.

Uma variável aleatória é uma função (mensurável) X que associa um número real a cada resultado de uma experiência aleatória. Consideremos uma aplicação com domínio Ω e conjunto de chegada \mathbb{R} ($X : \Omega \rightarrow \mathbb{R}$).

Em geral, temos:

$$\forall A \in F, \quad X(A) = \{X(w) : w \in A\}$$

$$\forall B \subset \mathbb{R}, \quad X^{-1}(B) = \{w \in \Omega : X(w) \in B\} \subset F$$

Definição 3.15. Seja (Ω, F, P) um espaço de probabilidades, uma *variável aleatória* é uma função:

$$X : \Omega \rightarrow \mathbb{R}$$

$$w \rightarrow X(w) \quad \text{tal que} \quad \{w : X(w) \leq x\} \in F, \forall x \in \mathbb{R}.$$

esquemáticamente temos:

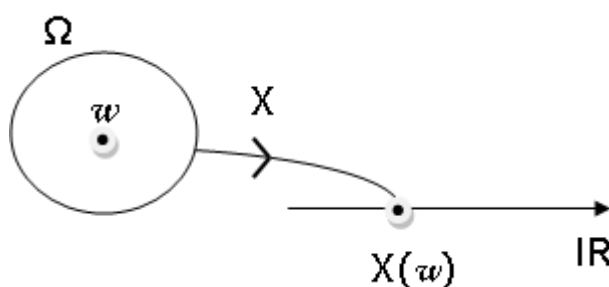


Figura 3.3: Representação esquemática de uma variável aleatória.

Se (Ω, F, P) for um espaço de probabilidade e X uma variável aleatória nele definido, ela induz uma distribuição de probabilidade em \mathbb{R} e temos:

$(\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B}, P_x)$ em que

- \mathbb{R} é o espaço de valores da v.a. X ;
- \mathcal{B} é σ -Álgebra de Borel (gerada por intervalos do tipo $]-\infty, x]$);
- P_x é a probabilidade induzida por X em \mathbb{R} e tal que

$$P_x(B) = P(X^{-1}(B)), \forall B \in \mathcal{B}.$$

$(\mathbb{R}, \mathcal{B}, P_x)$ é um espaço de probabilidade.

Exemplo 3.14.

Consideremos o lançamento de uma moeda. Seja X a variável definida no espaço amostral como o número de faces nacionais que se obtém em dois lançamentos (N - face nacional e C - face comum). O espaço amostral é $\Omega = \{CC, CN, NC, NN\}$ e a variável aleatória X toma os valores $\{0, 1, 2\}$.

Definição 3.16. Se um espaço amostral contém um número finito de acontecimentos ou uma sequência infinita numerável, ele é designado por *espaço amostral discreto*. A variável aleatória definida sobre esse espaço é designada por *variável aleatória discreta* (v.a.d.).

Assim, X é uma v.a.d. se é só se assume um número finito ou infinito numerável de valores distintos.

No exemplo anterior observamos que o número de elementos do espaço amostral é quatro e a variável pode tomar três valores distintos, logo X é uma v.a.d.

Exemplo 3.15.

Suponhamos que, após um exame médico, as pessoas foram diagnosticadas como tendo diabetes (D) ou não tendo diabetes (N). Admita que três pessoas sejam escolhidas ao acaso e classificadas de acordo com esse esquema.

O espaço amostral é dado por

$$\Omega = \{DDD, DDN, DND, NDD, NND, NDN, DNN, NNN\}.$$

O nosso interesse é saber quantas pessoas com diabetes foram encontradas, não interessando a ordem em que tenham sido selecionadas, ou seja, desejamos estudar a variável aleatória X , que a cada resultado, $\omega \in \Omega$, associa o número de pessoas com diabetes. Consequentemente, o conjunto de valores possíveis de X é $\{0, 1, 2, 3\}$ e assim, X é uma variável aleatória discreta.

Se um espaço amostral, Ω , é um conjunto infinito não numerável, a variável aleatória X pode tomar qualquer valor numérico num determinado intervalo ou coleção de intervalos, ou seja, a variável aleatória X assume valores dentro de intervalos de números reais, define-se assim que X é uma *variável aleatória contínua* (v.a.c.).

Exemplo 3.16. As variáveis aleatórias que representam o peso e a altura de uma população são contínuas.

No que se segue tratamos unicamente de variáveis aleatórias discretas que assumem um número finito de valores distintos.

Uma distribuição de probabilidade descreve as probabilidades que uma variável pode assumir ao longo de um espaço de valores. Ela é uma função cujo domínio são os valores da variável e cujo contradomínio são as probabilidades da variável assumir cada valor do domínio. Os elementos do contradomínio deste tipo de função só podem tomar valores entre 0 e 1.

Definição 3.17. Dada uma variável aleatória discreta X , que assume um número finito de valores distintos x_1, x_2, \dots, x_n , as probabilidades $p_i = P(X = x_i)$, $i = 1 \dots n$, devem satisfazer as seguintes condições:

$$(i) \quad 0 \leq p_i \leq 1, i = 1 \dots n;$$

$$(ii) \quad \sum_i P(x_i) = 1.$$

Designa-se função de probabilidade (f.p.) da variável X e denota-se por p_i , a função que a cada valor particular x_i associa a probabilidade de X ser igual a x_i . Simbolicamente,

$$p_i = P(x_i) = P(X = x_i).$$

Exemplo 3.17.

Consideremos novamente o exemplo em que uma moeda é lançada duas vezes. Seja X a variável definida no espaço amostral Ω que é igual ao número de faces nacionais que se obtém nos dois lançamentos (N - face nacional e C - face comum). A tabela seguinte apresenta a distribuição de probabilidade referente a variável aleatória X .

Valores de X	Acontecimentos	Probabilidades
0	(C,C)	$\frac{1}{4}$
1	(C,N), (N,C)	$\frac{1}{2}$
2	(N,N)	$\frac{1}{4}$

Tabela 3.2: Distribuição de probabilidade de X .

Os valores das probabilidades, apresentados na tabela 3.2, são obtidos da seguinte forma:

$$\begin{aligned}
 P(X = 0) &= P((C, C)) = \frac{1}{4}; \\
 P(X = 1) &= P((C, N)) + P((N, C)) = \frac{1}{2}; \\
 P(X = 2) &= P((N, N)) = \frac{1}{4}.
 \end{aligned}$$

Distribuição conjunta de duas variáveis aleatórias

Em relação a uma mesma experiência podem ser consideradas várias variáveis aleatórias. Vimos as distribuições de probabilidades para o caso de uma única variável aleatória discreta. Para duas ou mais v.a. o seu comportamento simultâneo é estudado usando as chamadas distribuições de probabilidade conjuntas. Assim sendo, a cada valor que (X, Y) pode assumir, deve ser associada a probabilidade de sua ocorrência. Precisamos, então, de definir a distribuição da v.a. bidimensional (X, Y) .

Definição 3.18. Seja (X, Y) uma v.a.d. bidimensional, em que os valores de X e de Y são finitos ou infinitos numeráveis. Assim, os valores de (X, Y) podem ser representados por (x_i, y_j) , $i = 1, 2, \dots, r$ e $j = 1, 2, \dots, s$. A função que a cada valor de (x_i, y_j) associa a

probabilidade da sua ocorrência, denota-se por $p(x_i, y_j) = P(X_i, Y_j) = P(X = x_i, Y = y_j)$, e designa-se por *função de probabilidade conjunta* se satisfizer as seguintes condições:

- $P(x_i, y_j) \geq 0$ para todo o (x_i, y_j) ;
- $\sum_i \sum_j P(x_i, y_j) = 1$.

Exemplo 3.18.

Consideremos o lançamento de dois dados perfeitos. Seja X o número de vezes que sai a face com o número seis, $x = 0, 1, 2$ e Y o número de vezes que sai a face com o número cinco, $y = 0, 1, 2$. Temos as seguintes probabilidades conjuntas:

$$P(X = 0, Y = 0) = \frac{4}{6} \times \frac{4}{6}.$$

$$P(X = 1, Y = 0) = \frac{4}{6} \times \frac{1}{6} \times 2 = P(X = 0, Y = 1).$$

$$P(X = 1, Y = 1) = \frac{1}{6} \times \frac{1}{6} \times 2.$$

$$P(X = 2, Y = 0) = \frac{1}{6} \times \frac{1}{6} = P(X = 0, Y = 2).$$

$$P(X = 1, Y = 2) = 0 = P(X = 2, Y = 1).$$

$$P(X = 2, Y = 2) = 0.$$

$$\sum_{x=0}^2 \sum_{y=0}^2 P(X = x, Y = y) = 1.$$

A seguir apresentamos a organização da função de probabilidade conjunta em tabela: A

$Y \backslash X$	0	1	2
0	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$
1	$\frac{8}{36}$	$\frac{2}{36}$	0
2	$\frac{1}{36}$	0	0

Tabela 3.3: Tabela da função de probabilidade conjunta do exemplo 3.18.

partir da tabela podemos calcular, também, as probabilidades para X ou Y , as probabili-

dades marginais. Por exemplo,

$$\begin{aligned} P(X = 0) &= P(X = 0, Y = 0) + P(X = 0, Y = 1) + P(X = 0, Y = 2) = \\ &= \sum_{y=0}^2 P(X = 0, Y = y) \end{aligned}$$

Estas probabilidades podem ser acrescentadas à tabela da função de probabilidade conjunta:

$Y \setminus X$	0	1	2	$P(Y = y)$
0	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$	$\frac{25}{36}$
1	$\frac{8}{36}$	$\frac{2}{36}$	0	$\frac{10}{36}$
2	$\frac{1}{36}$	0	0	$\frac{1}{36}$
$P(X = x)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	1

Tabela 3.4: Tabela da função de probabilidade conjunta completa do exemplo 3.18.

Definição 3.19. Seja X e Y v.a.d. de uma distribuição de probabilidade conjunta, podemos determinar a distribuição de probabilidade de X sem considerar Y e vice versa. Essas distribuições são designadas por *distribuições marginais*. A probabilidade marginal para cada valor é obtida da seguinte forma:

- para X : $P(X = x_i) = P(x_i) = \sum_{j=1}^s P(x_i, y_j)$;
- para Y : $P(Y = y_j) = P(y_j) = \sum_{i=1}^r P(x_i, y_j)$.

Distribuição condicionada: caso discreto [25]

Como, definido anteriormente, para dois acontecimentos quaisquer A e B , a probabilidade condicionada de A dado B é definida, desde que $P(B) > 0$, por $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Definição 3.20. Sejam X e Y variáveis aleatórias discretas e x_i um valor de X , tal que $P(X = x_i) = P(x_i) > 0$, define-se a *função de probabilidade condicionada* de $Y = y_j$ dado $X = x_i$, por

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}.$$

Assim, para x_i fixo, os pares $(y_j, P(Y = y_j, X = x_i))$ definem a *distribuição condicionada* de Y , dado que $X = x_i$.

Como se definiu, anteriormente, acontecimentos independentes, de modo análogo se define variáveis aleatórias independentes.

Definição 3.21. Seja (X, Y) v.a.d. bidimensional. Dizemos que X e Y são independentes se para todos os valores (x_i, y_j) de X e Y , tais que $P(X = x_i) > 0$ e $P(Y = y_j) > 0$:

$$P(X = x_i | Y = y_j) = P(X = x_i),$$

ou equivalentemente,

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j).$$

Exemplo 3.19.

Suponhamos que $P(X = x_i, Y = y_j)$, a função de probabilidade conjunta de X e Y , seja dada por $P(X = 0, Y = 0) = 0.4$, $P(X = 0, Y = 1) = 0.2$, $P(X = 1, Y = 0) = 0.1$, $P(X = 1, Y = 1) = 0.3$. Pretendemos determinar a probabilidade condicionada de X dado que $Y = 1$. Primeiro, observamos que

$$P(Y = 1) = \sum_i P(x_i, 1) = P(X = 0, Y = 1) + P(X = 1, Y = 1) = 0.5.$$

Então, $P(X = 0 | Y = 1) = \frac{P(X=0, Y=1)}{P(Y=1)} = \frac{2}{5}$ e $P(X = 1 | Y = 1) = \frac{P(X=1, Y=1)}{P(Y=1)} = \frac{3}{5}$.

Variáveis aleatórias multidimensionais

O estudo feito até agora, incidiu em variáveis aleatórias unidimensionais e bidimensionais, contudo é frequente termos interesse em estudar mais do que duas variáveis em simultâneo de uma mesma população. Nestas situações, queremos observar simultaneamente diversas características de uma população como por exemplo, a idade, X_1 , o número de filhos, X_2 e o número de carros de uma determinada família, X_3 . O terno ordenado (X_1, X_2, X_3) designa-se por *variável aleatória tridimensional*.

Definição 3.22. (X_1, X_2, \dots, X_n) é uma variável aleatória n -dimensional se $X_1 = x_1(s)$, $X_2 = x_2(s)$, \dots , $X_n = x_n(s)$ forem n funções, em que cada uma associa um número real a cada resultado s do espaço amostral S .

As componentes de um vetor aleatório são variáveis aleatórias que estão definidas sobre um espaço amostral comum. Tal como nos casos anteriores a variável aleatória pode ser discreta se todas as componentes do vetor forem discretos ou contínua se todas as componentes do vetor forem contínuas.

Definição 3.23. A distribuição de probabilidade conjunta de um conjunto de variáveis X_1, X_2, \dots, X_n é definida como $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$.

Findo esta breve revisão da teoria das probabilidades, em que conceitos fundamentais necessários para o entendimento das redes bayesianas, foram apresentados, passamos, no capítulo seguinte, para o estudo de redes bayesianas.

Capítulo 4

Redes Bayesianas

Neste capítulo apresentamos redes bayesianas, que permitem a representação gráfica da estrutura e modelação de um problema ou situação que envolve a incerteza através do raciocínio probabilístico. A teoria de probabilidades fornece meios para descrever e manipular o conhecimento incerto. Um de seus resultados mais úteis é o teorema de Bayes, que mostra uma maneira de calcular a probabilidade de um evento em particular, dado algum conjunto de observações que tenhamos feito.

Uma forma de estruturar uma situação para o raciocínio sob incerteza é construir um grafo que representa as relações causais entre eventos (variáveis) que são os nós do grafo e os arcos representam as relações de dependência probabilística (a influência que uma variável tem em relação à outra). A medida dessas relações é representada como probabilidades condicionadas, ou seja, a probabilidade de um evento acontecer, dado, ou não, a ocorrência de outro. Dada a influência entre variáveis a alteração do valor de alguma variável (evidência) será propagada pela rede, alterando o valor de outras. Essa propagação ocorre de acordo com um algoritmo que utiliza do teorema de Bayes. Assim, segundo Perotto [19] “A Rede Bayesiana pode ser vista, então, como uma “grande equação” que precisa manter-se verdadeira. Quando o valor de uma variável muda, as outras mudam adequadamente a fim de se adaptar e manter a coerência das relações”.

Dentro de um grafo acíclico orientado, existem três formas básicas de ligar duas variá-

veis, X e Y , através de uma terceira variável, Z , sendo que, esta terceira, pode ser filho de uma e pai da outra, pai das outras duas ou filho das outras duas e o grafo é construído a partir destas ligações que representam os tipos de conhecimento. Estas, classificam-se segundo Jensen et al. [9], como:

- Conexão serial.

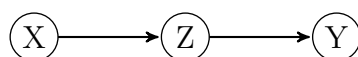


Figura 4.1: Conexão serial.

- Conexão divergente.

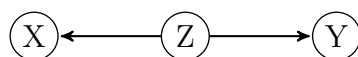


Figura 4.2: Conexão divergente.

- Conexão convergente.

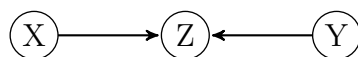


Figura 4.3: Conexão convergente.

Para cada uma destas possíveis ligações a propagação das evidências é estudada em [9] e [10], onde são definidos, com pormenor, esses três casos. São, ainda, apresentados vários exemplos que ilustram como as evidências podem ser transmitidas através de uma variável. Seguindo as regras é possível decidir para qualquer par de variáveis numa rede causal se elas são dependentes dadas as evidências conhecidas, ou seja, como a inserção de provas num determinado nó pode ou não ter influência na certeza doutros. Optámos por não desenvolver este estudo, dada a vastidão e complexidade da teoria de redes bayesianas, além de que, no modelo a descrever as evidências de um nó que recebe prova só são transmitidas para os seus nós pais e respetivos descendentes. Assim a nossa exposição sobre redes bayesianas incide na vertente mais ligada à teoria que sustenta a consistência

do modelo de aluno para redes bayesianas, em que se baseia a aplicação *web* a descrever neste trabalho.

Na literatura encontramos várias formas de definir redes bayesianas. Nesta dissertação optámos por usar a definição apresentada por Valldeperas em [17], uma vez que é a que melhor se adequa ao problema, que no nosso caso é a aplicação apresentada no capítulo 6. Em [9] e [10] podem ser encontradas outras formas de definir redes bayesianas.

Baseando-nos na tese de Eva [17], começaremos por apresentar uma noção intuitiva através de exemplos no contexto de ensino e passaremos para uma exposição mais formal onde as hipóteses de independência condicionada e o teorema de fatorização para redes bayesianas são apresentados, sendo estes as bases que tornam as redes bayesianas uma ferramenta muito poderosa para representar domínios com incerteza inerente.

Para a compreensão da teoria aqui apresentada sobre redes bayesianas, assume-se o conhecimento da teoria de grafos e do cálculo de probabilidades, conforme apresentado, respetivamente, nos capítulos 2 e 3 desta dissertação.

4.1 Apresentação Intuitiva

As redes bayesianas (*Bayesian Networks*) constituem modelos gráficos de representação das dependências entre variáveis de uma distribuição conjunta de probabilidade. Essa distribuição é modelada por um grafo orientado acíclico, no qual os nós representam as variáveis aleatórias da distribuição, que por sua vez representam uma entidade do mundo real, e os arcos representam as relações entre essas variáveis.

Apresentam-se, de seguida, exemplos no contexto deste trabalho.

Exemplo 4.1.

Vejamos o exemplo mais simples de rede bayesiana, composto por duas variáveis, que chamamos de C e A e que representam, respetivamente, o conhecimento de um determinado conceito e a capacidade de responder a uma pergunta, como se mostra na figura 4.4.

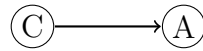


Figura 4.4: Exemplo de uma rede bayesiana.

O arco de C para A significa que conhecer o conceito C tem influência causal sobre o facto de responder corretamente ou não à pergunta A . O C é o *pai* de A e o A é o *filho* de C , como definido no capítulo 2.

A notação que usamos é a seguinte: Se X é uma variável binária, denotamos por $X = 1$ a presença e $X = 0$ a ausência da variável. Neste exemplo, $C = 1$, significa “o aluno sabe o conceito” e $C = 0$ “o aluno não sabe”; $A = 1$ significa “responde corretamente à questão A ” e $A = 0$ “responde incorretamente à questão”.

Vejamos um exemplo, na área da medicina, onde estes parâmetros são de simples interpretação:

Exemplo 4.2.

Suponhamos que temos uma rede que representa a relação entre ter uma determinada doença D ou não e o resultado um teste T utilizado para o diagnóstico da doença D .

A rede bayesiana é:

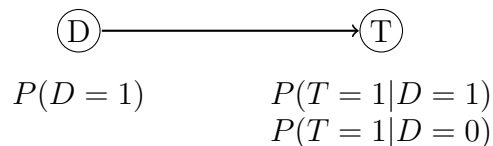


Figura 4.5: Exemplo de uma rede bayesiana em medicina.

Então:

- $P(D = 1)$ representa a probabilidade da população, em estudo, com a doença D , ou seja, a prevalência de D .
- $P(T = 1 | D = 1)$ indica a probabilidade de pacientes, de entre aqueles com a doença D , cujo teste T deu positivo. Esta probabilidade é conhecida como a *sensibilidade do teste*.

- O complementar da probabilidade anterior, $P(T = 0|D = 1) = 1 - P(T = 1|D = 1)$, representa a proporção de pacientes, de entre aqueles com a doença D , cujo teste deu negativo.
- $P(T = 1|D = 0)$ indica a probabilidade de pacientes, de entre os que não sofrem da doença D , cujo teste T deu positivo.
- O complementar da probabilidade anterior, $P(T = 0|D = 0)$, representa a proporção de pacientes, de entre aqueles sem a doença D , cujo teste T deu negativo. Esta probabilidade é designada por *especificidade do teste*.

Em medicina, procuram-se os testes com maior grau de sensibilidade e especificidade. Esta semântica pode ser estendida para o caso do modelo de aluno, onde falamos de sensibilidade e especificidade de uma questão para um conceito.

Voltando ao exemplo 4.1, a intensidade do relacionamento entre C e A é expressa pelas probabilidades condicionadas $P(A = 1|C = 1)$, $P(A = 1|C = 0)$, $P(A = 0|C = 1)$ e $P(A = 0|C = 0)$.

A informação quantitativa de uma rede Bayesiana é dada por probabilidades, sendo que estas podem ser classificadas de duas formas.

Definição 4.1. A *probabilidade à priori* é a probabilidade dos nós que não têm pais e a probabilidade condicionada dos nós que têm pais antes de obter qualquer evidência.

A *probabilidade à posteriori* é a probabilidade dos nós depois de obter evidências.

Exemplo 4.3.

No exemplo 4.1, os dados que devemos conhecer são os do nó C , que não tem pai, isto é $P(C = 1)$ e as probabilidades condicionadas do nó A , com pai, ou seja $P(A = 1|C = 1)$ e $P(A = 1|C = 0)$. Assim, a rede bayesiana completa é:

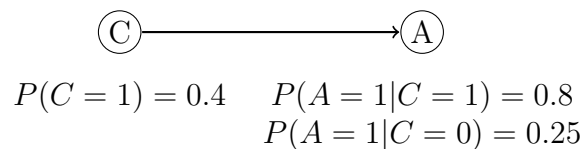


Figura 4.6: Rede bayesiana completa.

Podemos observar neste exemplo de rede bayesiana que:

- $P(C = 1) = 0,4$ indica que 40% dos alunos em estudo sabe o conceito.
- $P(A = 1|C = 1) = 0,8$ indica que 80% dos alunos que sabem o conceito C respondem corretamente à pergunta A . Isto quer dizer que os alunos que conhecem o conceito podem, ter um deslize e responder mal à pergunta numa proporção de 20%.
- $P(A = 1|C = 0) = 0,25$ significa que 25% dos alunos que não sabe o conceito C são capazes de responder corretamente à pergunta A . Este parâmetro indica, também, que alunos que não sabem o conceito podem adivinhar a resposta correta à pergunta A .

Estas relações são apresentadas na tabela de probabilidades conjuntas e tabela de probabilidades condicionadas (TPC), que listam as probabilidades de cada um dos nós filhos em relação às diferentes combinações de valores dos nós pais.

	$C = 0$	$C = 1$
$A = 0$	$P(A = 0 \cap C = 0)$ $= P(A = 0 C = 0)P(C = 0)$	$P(A = 0 \cap C = 1)$ $= P(A = 0 C = 1)P(C = 1)$
$A = 1$	$P(A = 1 \cap C = 0)$ $= P(A = 1 C = 0)P(C = 0)$	$P(A = 1 \cap C = 1)$ $= P(A = 1 C = 1)P(C = 1)$

Tabela 4.1: Tabela de probabilidades conjuntas.

	$C = 0$	$C = 1$
$A = 0$	$P(A = 0 C = 0)$	$P(A = 0 C = 1)$
$A = 1$	$P(A = 1 C = 0)$	$P(A = 1 C = 1)$

Tabela 4.2: Tabela de probabilidades Condicionadas.

Conhecendo estes dados, podemos calcular:

a) a probabilidade à priori de que um aluno responda corretamente à pergunta A.

$$\begin{aligned}
 P(A = 1) &= P(A = 1 \cap C = 1) + P(A = 1 \cap C = 0) \\
 &= P(A = 1|C = 1)P(C = 1) + P(A = 1|C = 0)P(C = 0) \\
 &= 0.8 \times 0.4 + 0.25 \times 0.6 \\
 &= 0.47.
 \end{aligned}$$

$$P(A = 0) = 1 - P(A = 1) = 1 - 0.47 = 0.53.$$

ou

$$\begin{aligned}
 P(A = 0) &= P(A = 0 \cap C = 1) + P(A = 0 \cap C = 0) \\
 &= P(A = 0|C = 1)P(C = 1) + P(A = 0|C = 0)P(C = 0) \\
 &= (1 - P(A = 1|C = 1))P(C = 1) + (1 - P(A = 1|C = 0))P(C = 0) \\
 &= 0.2 \times 0.4 + 0.75 \times 0.6 \\
 &= 0.53.
 \end{aligned}$$

Resumidamente:

	$C = 0$	$C = 1$	Total
$A = 0$	0.75×0.6	0.2×0.4	0.53
$A = 1$	0.25×0.6	0.8×0.4	0.4
Total	0.4	0.6	1

b) as probabilidades à posteriori dada uma evidência observada E ,

$$P^*(C) = P(C|E).$$

Suponhamos que a evidência observada é que certo aluno respondeu corretamente à pergunta A. Pretendemos determinar a probabilidade de que esse aluno conheça o conceito C.

Se não existisse a possibilidade de um aluno que não sabe o conceito C responder bem à pergunta A , $P(C) = 1$, mas como assim não é, temos de calcular

$$P^*(C = 1) = P(C = 1|A = 1).$$

Aplicando o teorema de Bayes, obtemos:

$$\begin{aligned} P^*(C = 1) &= P(C = 1|A = 1) \\ &= \frac{P(A = 1|C = 1)P(C = 1)}{P(A = 1)} \\ &= \frac{0.8 \times 0.4}{0.47} \\ &= 0.68. \end{aligned}$$

Da mesma forma podemos calcular $P^*(C = 0)$:

$$\begin{aligned} P^*(C = 0) &= P(C = 0|A = 1) \\ &= \frac{P(A = 1|C = 0)P(C = 0)}{P(A = 1)} \\ &= \frac{0.25 \times 0.6}{0.47} \\ &= 0.32. \end{aligned}$$

ou, por ser a probabilidade complementar,

$$P(C = 0) = 1 - P(C = 1) = 1 - 0.68 = 0.32.$$

Suponhamos, agora, que a evidência observada é que certo aluno respondeu incorretamente à pergunta A . A probabilidade à posteriori de que o aluno conheça o conceito C calcula-se com um procedimento totalmente análogo ao anterior.

Exemplo 4.4. Suponhamos que ampliamos o modelo anterior acrescentando outra pergunta, B . A rede bayesiana mostra-se na Figura 4.7.

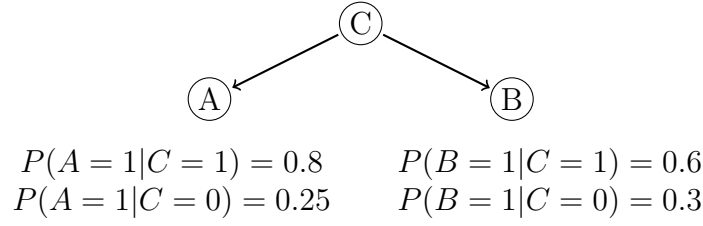


Figura 4.7: Rede bayesiana do exemplo 4.4.

Notamos que a questão B é menos sensível e menos específica para o conceito C que a pergunta A . Pois, dado que sabe o conceito C , a probabilidade de acertar na questão B é menor do que a probabilidade de acertar na questão A , o que significa que B traduz menor sensibilidade. Além disso, dado que o aluno não sabe o conceito C , a probabilidade de acertar em B é maior do que a de acertar em A , o que significa B traduz menor especificidade para o conceito C . Vejamos a tabela de probabilidades conjuntas e veremos que tipo de conclusões podemos tirar desta informação.

	A		B	
	$A = 0$	$A = 1$	$B = 0$	$B = 1$
$C = 0$	$P(A = 0 C = 0)P(C = 0)$ $= 0.75 \times 0.6 = 0.45$	$P(A = 1 C = 0)P(C = 0)$ $= 0.25 \times 0.6 = 0.15$	$P(B = 0 C = 0)P(C = 0)$ $= 0.7 \times 0.6 = 0.42$	$P(B = 1 C = 0)P(C = 0)$ $= 0.3 \times 0.6 = 0.18$
$C = 1$	$P(A = 0 C = 1)P(C = 1)$ $= 0.2 \times 0.4 = 0.08$	$P(A = 1 C = 1)P(C = 1)$ $= 0.8 \times 0.4 = 0.32$	$P(B = 0 C = 1)P(C = 1)$ $= 0.4 \times 0.4 = 0.16$	$P(B = 1 C = 1)P(C = 1)$ $= 0.6 \times 0.4 = 0.24$

Tabela 4.3: Tabela de probabilidades conjuntas.

a) Suponha que o aluno responde à pergunta B corretamente, ou seja, que a evidência disponível é $E = \{B = 1\}$. Então, podemos calcular, como atrás, a probabilidade à posterior de que o aluno conhece o conceito C , uma vez que respondeu bem a questão B :

$$\begin{aligned}
 P^*(C = 1) &= P(C = 1|B = 1) \\
 &= \frac{P(B = 1|C = 1)P(C = 1)}{P(B = 1)} \\
 &= \frac{0.6 \times 0.4}{P(B = 1|C = 1)P(C = 1) + P(B = 1|C = 0)P(C = 0)} = \\
 &= \frac{0.6 \times 0.4}{0.6 \times 0.4 + 0.3 \times 0.6} \\
 &\simeq 0.57.
 \end{aligned}$$

Como podemos ver, a probabilidade de que um estudante que responda bem à pergunta B conheça o conceito C é 0,57, ligeiramente menor do que o resultado, no caso em que responda bem a A (0,68), o que é explicado pela menor sensibilidade e especificidade da questão B .

b) Suponhamos agora que a evidência (E) observada é que o aluno responde corretamente às duas perguntas $E = \{A = 1 \cap B = 1\}$. Pretendemos calcular a probabilidade de saber o conceito C , dado que acertou nas duas perguntas. Usando, novamente teorema de Bayes, temos:

$$P^*(C = 1) = P(C = 1|A = 1, B = 1) = \frac{P(A = 1, B = 1|C = 1)P(C = 1)}{P(A = 1, B = 1)}. \quad (4.1)$$

Para podermos determinar $P(A = 1, B = 1|C = 1)$ usamos a fatorização de probabilidades apresentada mais à frente neste texto. Especificamente assume-se que $P(A = 1, B = 1, C = 1) = P(C = 1)P(A = 1|C = 1)P(B = 1|C = 1)$ dado que C é o seu pai comum de A e B na rede. Assim temos que

$$\begin{aligned} P(A = 1, B = 1|C = 1) &= \frac{P(A = 1, B = 1, C = 1)}{P(C)} \\ &= \frac{P(C = 1)P(A = 1|C = 1)P(B = 1|C = 1)}{P(C)} \\ &= P(A = 1|C = 1)P(B = 1|C = 1). \end{aligned}$$

Isto significa que assumir que um aluno que sabe o conceito C , o facto de acertar ou errar a pergunta B não depende quer de acertar ou de errar a pergunta A . Supondo isto temos então:

$$P(A = 1, B = 1|C = 1) = P(A = 1|C = 1) \times P(B = 1|C = 1) = 0.8 \times 0.6 = 0.48.$$

$$\begin{aligned}
P(A = 1, B = 1) &= P(A = 1, B = 1|C = 1)P(C = 1) + \\
&\quad + P(A = 1, B = 1|C = 0)P(C = 0) \\
&= P(A = 1|C = 1) \times P(B = 1|C = 1)P(C = 1) + \\
&\quad + P(A = 1|C = 0) \times P(B = 1|C = 0)P(C = 0) \\
&= 0.8 \times 0.6 \times 0.4 + 0.25 \times 0.3 \times 0.6 \\
&= 0.237.
\end{aligned}$$

Substituindo em 4.1, vem

$$P^*(C = 1) = \frac{0.48 \times 0.4}{0.237} = 0.81.$$

Assim:

$$P^*(C = 1) = 0.81.$$

$$P^*(C = 0) = 0.19.$$

Como esperado, quando temos duas evidências em favor do aluno que conhece o conceito, a probabilidade resultante é maior do que a correspondente a cada uma delas separadamente.

4.2 Apresentação Formal

Matematicamente, uma rede bayesiana é uma representação compacta de uma tabela de probabilidades conjuntas do universo dum problema. Por outro lado, do ponto de vista de um especialista, redes bayesianas constituem um modelo gráfico que representa de forma simples as relações de causalidade das variáveis de um sistema, ou seja, consiste num conjunto de variáveis e num conjunto de arcos orientados entre as variáveis. Matematicamente, a estrutura é chamado um grafo orientado. Ao falar sobre as relações num grafo orientado usamos a expressão das relações familiares: se há uma ligação de X para Y, dizemos que Y é filho de X, e X é um pai de Y, como no exemplo seguinte:

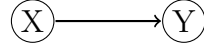


Figura 4.8: Rede bayesiana com duas variáveis.

Uma variável pode ter qualquer número de estados, pode, por exemplo, ser a cor de olhos (estados: azul, verde, preto, castanho), o número de filhos de um casal (estados: 0, 1, 2, 3, 4, 5, 6, maior do que 6), ou uma doença (estados: sida, tuberculose, cancro de pulmão). O nosso estudo vai incidir apenas sobre as variáveis com estado binário, pois é este o tipo de variável definida no modelo do aluno da aplicação a estudar no próximo capítulo. Baseando em [17], apresentamos, de seguida, a definição formal de redes bayesianas.

Definição 4.2. Uma rede bayesiana é constituída por

- um conjunto de variáveis proposicionais V ;
- um conjunto, E , de relações binárias definidas sobre as variáveis V , ou seja um conjunto de arestas orientadas entre as variáveis e em que cada variável tem um conjunto finito de estados mutuamente exclusivos;
- uma distribuição de probabilidade conjunta P definida para as variáveis de V , de tal modo que $G = (V, E)$ é um grafo acíclico, conexo e orientado e (G, P) satisfaz a hipótese de independência condicionada, também chamada de separação direcional, que se define de seguida.

Definição 4.3. Seja $G = (V, E)$ um grafo acíclico conexo e orientado e seja P uma distribuição de probabilidade conjunta definida sobre as variáveis do grafo. Diz-se que as variáveis satisfazem as *hipóteses de independência condicionada* (HIC), se para cada variável X de V , o conjunto de pais diretos de X separa condicionalmente X de todo o subconjunto Y de V que não contém X ou algum de seus descendentes. Ou seja,

$$\forall X \in V, \forall Y \subseteq V \setminus \{X \cup de(X)\} \quad \text{tem-se} \quad P(X|pa(X), Y) = P(X|pa(X)). \quad (4.2)$$

Exemplo 4.5. Consideremos o seguinte exemplo de rede bayesiana:

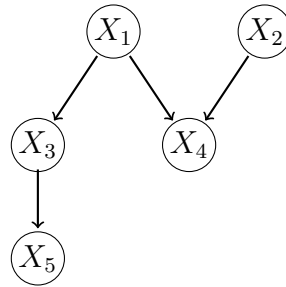


Figura 4.9: Rede bayesiana com cinco variáveis.

Seja $Y = \{X_1, X_2, X_4\}$ um subconjunto do conjunto das variáveis que constituem a rede, a que não pertence a variável X_3 nem os descendentes. Note-se que $pa(X_3) = \{X_1\}$ e, pela HIC, temos que $P(X_3|X_1, X_2, X_4) = P(X_3|X_1)$.

Na definição da rede bayesiana, começamos a partir de uma distribuição de probabilidade conjunta para as variáveis. Para definir uma rede bayesiana com n nós, em que cada nó representa uma variável binária, temos de saber $2^n - 1$ valores de probabilidades. No entanto, as hipóteses de independência condicionada permitem que não seja necessário conhecer todos esses valores, uma vez que, como veremos no teorema que se segue, a distribuição de probabilidade conjunta pode ser expressa como o produto das distribuições condicionadas de cada nó dado os seus pais.

Teorema 4.1. *Fatorização de probabilidade*

Dada uma rede bayesiana, a distribuição de probabilidade conjunta pode ser determinada da seguinte forma:

$$P(X_1, \dots, X_n) = \prod_i P(X_i | pa(X_i)). \quad (4.3)$$

Demonstração.

É fácil construir uma ordem de variáveis em que os pais de cada nó aparecem sempre atrás dele. Suponhamos, portanto, que a ordenação de X_1, \dots, X_n satisfaz essa propriedade. Uma vez que, $P(X_1, \dots, X_n) = \prod_i P(X_i | X_{i+1}, \dots, X_n)$, pela fórmula das probabilidades compostas, e ao escolher a ordenação, o conjunto $\{X_{i+1}, \dots, X_n\}$ inclui todos os pais de X_i ,

usando a HIC temos $P(X_i|X_{i+1}, \dots, X_n) = \prod_i P(X_i|pa(X_i))$, pois o conjunto $\{X_{i+1}, \dots, X_n\}$ não contam X_i nem descendentes de X_i . \square

Exemplo 4.6.

No exemplo 4.5 em que $n = 5$, teríamos de saber $2^5 - 1 = 32 - 1 = 31$ valores. Mas com as condições HIC temos 10 probabilidades, sendo elas

$P(X_1 = 1)$, $P(X_2 = 1)$, $P(X_3 = 1|X_1 = 1)$, $P(X_3 = 1|X_1 = 0)$, $P(X_4 = 1|X_1 = 1, X_2 = 1)$, $P(X_4 = 1|X_1 = 0, X_2 = 1)$, $P(X_4 = 1|X_1 = 1, X_2 = 0)$, $P(X_4 = 1|X_1 = 0, X_2 = 0)$, $P(X_5 = 1|X_3 = 1)$ e $P(X_5 = 1|X_3 = 0)$.

Teremos então que definir empiricamente para cada variável X_i com pais X_{i+1}, \dots, X_n , uma probabilidade condicionada $P(X_i|X_{i+1}, \dots, X_n)$. Note-se que, se X_i não tem pais, então a probabilidade condicionada reduz-se à probabilidade $P(X_i)$.

Exemplo 4.7.

Considere a situação da figura 4.10

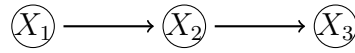


Figura 4.10: Rede bayesiana com três variáveis.

Assumindo que todas as variáveis são binárias, para este exemplo de rede bayesiana, caso em que $n = 3$, seriam necessários $2^3 - 1 = 7$, sendo elas:

$P(X_1 = 1)$; $P(X_2 = 1|X_1 = 1)$; $P(X_2 = 1|X_1 = 0)$; $P(X_3 = 1|X_1 = 1, X_2 = 1)$;
 $P(X_3 = 1|X_1 = 0, X_2 = 1)$; $P(X_3 = 1|X_1 = 1, X_2 = 0)$; $P(X_3 = 1|X_1 = 0, X_2 = 0)$.

Tendo em conta as condições de independência condicional teremos $2 \times 3 - 1 = 5$, sendo elas $P(X_1 = 1)$; $P(X_2 = 1|X_1 = 1)$; $P(X_2 = 1|X_1 = 0)$; $P(X_3 = 1|X_2 = 1)$; $P(X_3 = 1|X_2 = 0)$.

E, além de consultas a partir de evidências, redes Bayesianas podem ser utilizadas para:

- Tomar decisões baseadas em probabilidades.
- Decidir quais evidências adicionais devem ser observadas a fim de se obter informações úteis do sistema.

- Analisar o sistema a fim de buscar os aspetos do modelo que possuem maior impacto sob as variáveis de consulta.
- Explicar os resultados de uma inferência probabilística ao usuário.

No capítulo seguinte particularizamos o estudo de redes bayesianas um modelo genérico do aluno que servirá como base para o entendimento da aplicação *web* apresentar no capítulo seis.

Capítulo 5

Modelo genérico do aluno baseado em redes bayesianas

Como já referido, uma rede bayesiana modela uma situação real, em que o seu objetivo central (enfoque) é reter as informações que lhe são submetidas através de factos, a que chamamos evidências E^1 . Uma das tarefas mais importantes consiste em obter estimativas de probabilidades de eventos relacionados com os dados, à medida que novas informações ou evidências sejam conhecidas. Tendo portanto, como tarefa básica de inferência probabilística calcular a distribuição de probabilidades à posteriori para um conjunto de variáveis de consulta (o que queremos saber, no caso da aplicação a estudar é o nível de conhecimento dos conceitos pelos alunos) dadas as evidências (que no caso da aplicação a estudar são as respostas às questões). A principal vantagem em usar-se redes bayesianas está relacionada a sua capacidade de obter informações implícitas de informações explícitas, declaradas num grafo.

Neste capítulo vamos apresentar um breve resumo das principais características do modelo genérico do aluno baseado em redes bayesianas (MGARB). Para um estudo mais detalhado podemos consultar Millán & Pérez de la Cruz [15] e para saber como construir modelos de alunos baseados em redes bayesianas podemos consultar Millán, Loboda e Pérez

¹Evidência de uma variável é uma declaração de uma certeza do seu estado.

-de-la- Cruz [16]. Começamos por descrever os diferentes tipos de variáveis que fazem parte do modelo do aluno, a definição da sua estrutura, ou seja, os tipos de nós e as relações causais a estabelecer entre eles e por fim apresentamos os parâmetros a especificar para os diferentes tipos de nós na rede, bem como uma forma de simplificar esses parâmetros, dado que o seu número é bastante elevado. Os conceitos apresentados são baseados na tese de doutoramento de Eva Millán [17] e o artigo de Castillo et al. [4].

5.1 Variáveis definição

Existem dois tipos de variáveis que fazem parte do modelo do aluno, variáveis cujo objetivo é medir o conhecimento dos alunos e variáveis utilizadas para mostrar ou provar esses conhecimentos. Assim, o MGARB é composto por dois tipos de variáveis: as variáveis conhecimento e as variáveis de prova.

Variáveis de conhecimento representam o conhecimento dos alunos. Estas são as variáveis de interesse em sistemas de aprendizagem adaptativa, pois serão capazes de se adaptar a cada estudante individualmente e os seus valores não são diretamente observáveis. Estas variáveis podem medir conceitos simples que não englobam outros conceitos ou conceitos compostos que resultam da agregação de outras variáveis (subconceitos). O nível de detalhe necessário depende da natureza da aplicação.

De seguida, apresentamos um exemplo deste modelo, o qual servirá como base na aplicação a descrever.

Um domínio de conhecimento pode ser decomposto hierarquicamente em três níveis:

- i. Nível inferior - conceitos Base (C).
- ii. Nível médio - Temas (T) que representam os conceitos agregados.
- iii. Nível superior - Assunto (A) representam os temas agregados.

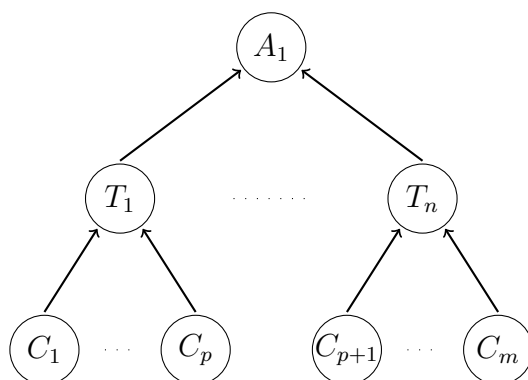


Figura 5.1: Uma estrutura do modelo de aluno com nós conhecimento.

No MGARB, todas as variáveis de conhecimento são modeladas como binárias: 0 (não sabe) e 1 (sabe).

Variáveis de prova, representam as ações dos alunos, e são diretamente observáveis, como por exemplo os resultados da resposta a uma questão de escolha múltipla. Os valores de tais variáveis são utilizadas para inferir os valores das variáveis de conhecimento oculto. No MGARB, as variáveis de prova, também se consideram binárias, com valor 0 (incorreto) ou 1 (correto).

5.2 Definição da estrutura

O próximo passo é definir as relações causais entre eles. As relações modeladas no MGARB são:

- Relacionamentos de agregação entre os nós de conhecimento (conceitos básicos, temas e assuntos): a MGARB considera que o conhecimento das partes tem uma influência causal sobre o conhecimento do assunto mais geral.
- As relações entre os nós de conhecimento e os nós de prova: a MGARB considera que ter conhecimento sobre determinados conceitos tem influência causal sobre ser capaz de resolver corretamente os exercícios associados.

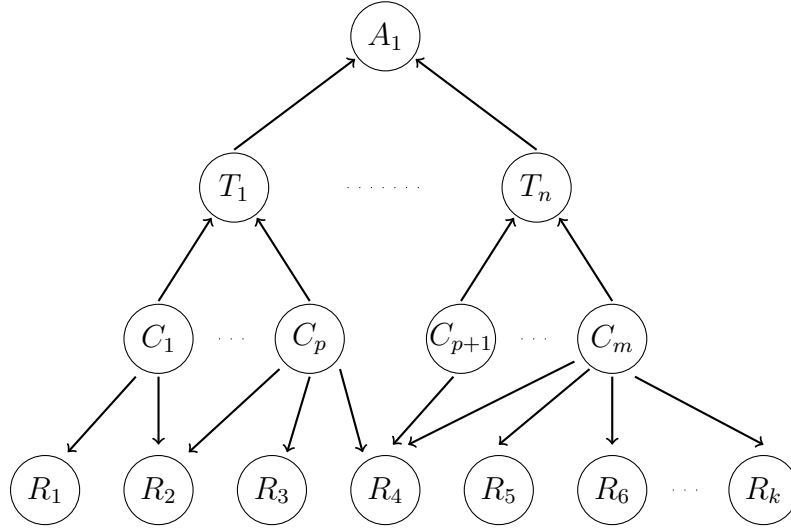


Figura 5.2: A estrutura do MGARB.

5.3 Estimativa de parâmetros

Os parâmetros necessários a especificar para este exemplo do modelo de aluno são:

5.3.1 Estimativa de parâmetros para nós que representam conceitos básicos

Para cada conceito básico, C_i , $i \in 1, \dots, m$, é necessário definir o valor da sua probabilidade à priori $P(C_i)$. Como referido anteriormente temos que C_i é uma variável binária, ou seja, $C_i = 0$ ou $C_i = 1$, assim, a probabilidade de C_i é dada por: $P(C_i = x) = p_i^x(1 - p_i)^{1-x}$ onde p_i é a probabilidade de que o aluno conheça o conceito C_i , $i \in \{1, \dots, m\}$, e x pode tomar o valor 0 ou 1.

Para cada conceito C_i , p_i pode ser estimado a partir dos dados disponíveis, se existirem, ou caso contrário, ajustado uniformemente.

Exemplo 5.1.

Consideremos que $p_1 = 0.75$, isto significa que $P(C_1 = 1) = 0.75^1 \times (1 - 0.75)^0 = 0.75$ e $P(C_1 = 0) = 0.75^0 \times (1 - 0.75)^1 = 0.25$.

5.3.2 Estimativa de parâmetros para nós que representam temas

Considerando a rede bayesiana da figura 5.3. Para cada variável tema T é necessário definir as correspondentes probabilidades condicionadas dados os conceitos que o tema envolve. Isto é feito utilizando um vetor de pesos que define a importância de cada conceito no tema.

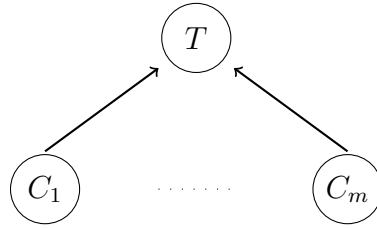


Figura 5.3: Excerto da rede da figura 5.2.

Seja $W = (w_1, w_2, \dots, w_m)$ um vetor de pesos que mede a importância de cada conceito C_i , no tema T , tal que $w_i > 0$, $i = 1, \dots, m$ e $\sum_{i=1}^m w_i = 1$.

As probabilidades condicionadas dos nós temas (filhos) dado os respectivos conceitos agregados (pais) são calculadas da seguinte forma:

$$P(T = 1 | (\{C_i = 1, i \in I\}, \{C_i = 0, i \notin I\})) = \sum_{i \in I} w_i$$

onde $I = \{i \in \{1, 2, \dots, m\} \text{ tal que } C_i = 1\}$.

Exemplo 5.2.

Considere os nós da Figura 5.4 que ilustra como o grau de conhecimento de um conceito influencia diretamente as probabilidades de um estudante saber um determinado tema. Nela, pretende-se estabelecer a influência causal das variáveis conceito “derivadas-D” e “integrais-I”, sobre o conceito “cálculo-C”.

O conjunto de variáveis $V = \{C, D, I\}$ retrata as variáveis do modelo; os círculos representam tanto os nós da rede quanto as variáveis do conjunto V , sendo estas binárias; os arcos representam o relacionamento causal entre as variáveis; e os parâmetros numéricos são representados pelas distribuições condicionadas do modelo. Supõe-se que, neste exemplo,

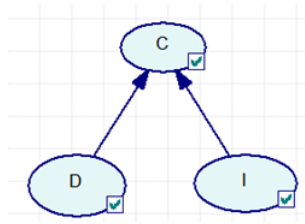


Figura 5.4: Rede bayesiana com variáveis conhecimento.

	$D = 1$		$D = 0$	
	$I = 1$	$I = 0$	$I = 1$	$I = 0$
$C = 1$	$1 \times w_1 + 1 \times w_2$ $= 0.6 + 0.4 = 1$	$1 \times w_1 + 0 \times w_2$ $= 0.6 + 0 = 0.6$	$0 \times w_1 + 1 \times w_2$ $= 0 + 0.4 = 0.4$	$0 \times w_1 + 0 \times w_2$ $= 0$

Tabela 5.1: Distribuições condicionadas do modelo MGARB.

$p_1 = p_2 = 0.5$ e $w_1 = 0.6$ e $w_2 = 0.4$. Estas ponderações w_1 e w_2 , são utilizadas para definir a tabela de probabilidade condicional do nó C dado os seus pais D e I .

Pela leitura da tabela concluímos que:

- $P(D = 1) = 0.5; P(I = 1) = 0.5$.
- $P(C = 1|D = 1, I = 1) = 1; P(C = 1|D = 1, I = 0) = 0.6$.
- $P(C = 1|D = 0, I = 1) = 0.4; P(C = 1|D = 0, I = 0) = 0$.

A distribuição conjunta das variáveis do modelo pode ser expressa via

$$P(C, D, I) = P(D)P(I)P(C|D, I),$$

lembrando que os conceitos D e I influenciam o conhecimento do conceito C .

Suponhamos que se queria conhecer a probabilidade de saber o conceito D dadas as observações dos outros conceitos, esta probabilidade não está armazenada diretamente no modelo e é necessário determiná-la. O cálculo das probabilidades pode ser realizado mediante uma simplificação e fatorização de variáveis. Por exemplo, no caso em que

a observação é a de que sabe os conceitos “cálculo” e “integrais” e pretende inferir a probabilidade de saber “derivadas”:

$$P(D = 1|C = 1, I = 1) = \frac{P(D = 1, C = 1, I = 1)}{P(C = 1, I = 1)}.$$

Dadas as independências condicionais a equação acima pode ser expressa como:

$$\begin{aligned} P(D = 1|C = 1, I = 1) &= \frac{P(D = 1)P(I = 1)P(C = 1|D = 1, I = 1)}{P(C = 1, I = 1, D = 1) + P(C = 1, I = 1, D = 0)} \\ &= \frac{0.5 \times 0.5 \times 1}{0.5 \times 0.5 \times 1 + P(D = 0)P(I = 1)P(C = 1|D = 0, I = 1)} \\ &= \frac{0.25}{0.25 + 0.5 \times 0.5 \times 0.4} \\ &= 0.714. \end{aligned}$$

Exemplo 5.3.

Suponhamos, agora, que à rede do exemplo 5.2 adicionamos um nó que representa “equações diferenciais” em que definimos os pesos w_1 , w_2 e w_3 , respetivamente, de cada conceito pai, nós D , I e E , que influenciam o conceito filho, C , como se mostra na figura seguinte:

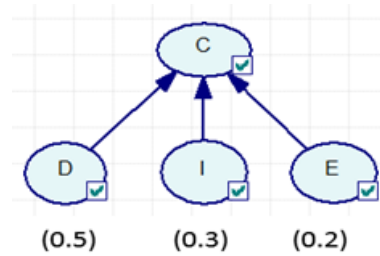


Figura 5.5: Rede bayesiana de conceitos.

Como podemos verificar a soma dos pesos é igual a 1.

A tabela de probabilidade condicionada para o nó C , isto é, para o conhecimento do conceito cálculo dado o conhecimento dos conceitos, derivada (D), integrais (I) e equações diferenciais (E), seria a seguinte:

P(C=1								
D=	1				0			
I=	1		0		1		0	
E=	1	0	1	0	1	0	1	0
J=	1	0.8	0.7	0.5	0.5	0.3	0.2	0

Figura 5.6: Tabela de probabilidades condicionada do exemplo 5.3.

Nesta tabela constam $2^3 = 8$ probabilidades associadas ao nó C . Por exemplo, temos que $P(C = 1|D = 1, I = 0, E = 1) = 0.7$ que se determinou da seguinte forma

$$\begin{aligned}
 P(C = 1|D = 1, I = 0, E = 1) &= 1 \times w_1 + 0 \times w_2 + 1 \times w_3 \\
 &= 1 \times 0.5 + 0 \times 0.3 + 1 \times 0.2 \\
 &= 0.7.
 \end{aligned}$$

Como apresentado nos exemplos anteriores, a tarefa fundamental para qualquer sistema de inferência probabilística é calcular a distribuição de probabilidade à posteriori para um conjunto de variáveis, dado o valor exato de alguma outra variável. Russell e Norving [22] denominam as primeiras “variáveis consulta” e as últimas “variáveis evidência”. Deste modo o sistema calcula $P(\text{Consulta}|\text{Evidencia})$. As redes bayesianas são flexíveis a ponto de qualquer nó poder servir como “consulta” ou como “evidência”.

5.3.3 Estimativa de parâmetros para nós que representam evidências

Para cada variável prova R_i , $i \in \{1, \dots, k\}$, ou seja, para cada nó com evidência dada pelo aluno é necessário definir as probabilidades condicionadas dos nós prova, dado os conceitos envolvidos C_i , $i \in \{1, \dots, m\}$, $P(R_i|C_1, \dots, C_m)$. Com estes nós pretendemos recolher informações sobre o estado do conhecimento dos alunos, sendo que estas informações podem vir de vários tipos de fontes, como por exemplo, de questões de escolha múltipla, em que o aluno tem de escolher uma só resposta entre quatro das apresentadas, pois só uma está correta. Neste tipo de nós também as variáveis são binárias tomando os valores

1 e 0, para o caso de acertarem ou errarem, respetivamente, a resposta à questão. De igual modo se define a probabilidade para o caso em que as questões sejam de desenvolvimento, contudo só é possível este tipo de perguntas se o sistema tiver a capacidade de comprovar se a solução dada pelo aluno está correta ou não. Neste tipo de redes bayesianas os nós evidência podem ser usados mais do que uma vez, ou seja, têm a capacidade de atualizar a evidência conhecida antes pela depois, o que se vai propagar por toda a árvore de conceitos e desta forma todas as probabilidades na rede são atualizadas. Este mecanismo de atualização é a aplicação direta da lei de probabilidade total e da probabilidade condicionada, como se mostra no exemplo que se segue.

Exemplo 5.4.

Suponhamos, agora, para o mesmo exemplo 5.2, anterior, que o aluno deu uma resposta relativa ao conceito D (evidência sobre D). Precisamos de definir à priori as seguintes probabilidades.

$$P(R = 1|D = 1) = 0.8 \text{ e } P(R = 1|D = 0) = 0.25.$$

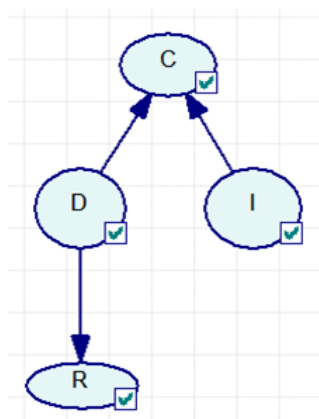


Figura 5.7: Rede bayesiana com evidência.

A probabilidade à posteriori de o aluno saber o conceito, dado que acertou na pergunta, isto é, $P(D = 1|R = 1)$, determina-se da seguinte forma:

$$\begin{aligned}
 P(D = 1|R = 1) &= \frac{P(R = 1|D = 1)P(D = 1)}{P(R = 1)} \\
 &= \frac{P(R = 1|D = 1)P(D = 1)}{P(R = 1, D = 1) + P(R = 1, D = 0)} \\
 &= \frac{P(R = 1|D = 1)P(D = 1)}{P(R = 1|D = 1)P(D = 1) + P(R = 1|D = 0)P(D = 0)} \\
 &= \frac{0.8 \times 0.5}{0.8 \times 0.5 + 0.25 \times 0.5} \\
 &= 0.76.
 \end{aligned}$$

A fatorização da distribuição de probabilidades representada pelo grafo é

$$\begin{aligned}
 P(C = c, D = d, I = i, R = r) \\
 = P(D = d)P(I = i)P(R = r|D = d)P(C = c|D = d, I = i).
 \end{aligned}$$

Para determinarmos, por exemplo, a probabilidade de saber o conceito “cálculo” dado que acertou na resposta R , temos de efetuar os cálculos que a seguir se apresentam, utilizando a fatorização de probabilidades e a regra de Bayes (condicionada e a probabilidade total).

$$\begin{aligned}
 P(C = 1|R = 1) &= \frac{P(C = 1, R = 1)}{P(R = 1)} \\
 &= \frac{\sum_{d,i} P(C = 1, R = 1, D = d, I = i)}{\sum_d P(R = 1, D = d)} \\
 &= \frac{\sum_{d,i} P(D = d)P(I = i)P(R = 1|D = d)P(C = 1|D = d, I = i)}{\sum_d P(R = 1|D = d)P(D = d)} \\
 &= 0.657.
 \end{aligned}$$

Consideremos que adicionamos uma evidência relativa aos conceitos D e I , na rede do exemplo 5.3, a rede agora seria:

Teremos de definir as probabilidades condicionadas de R dado os seus nós pai, D e I , como por exemplo se apresenta na tabela seguinte:

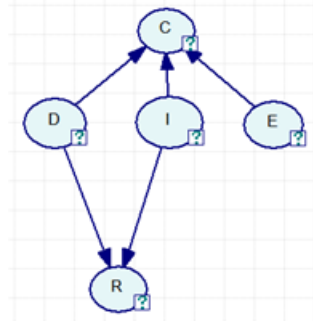


Figura 5.8: Rede bayesiana - propagação.

P(R=1)				
D=	1		0	
I=	1	0	1	0
P=	0.8	0.6	0.4	0.25

 Figura 5.9: Tabela de probabilidades condicionadas associada ao nó da resposta R .

A fatorização da distribuição de probabilidades representada pelo grafo é

$$\begin{aligned}
 P(C = c, D = d, I = i, E = e, R = r) \\
 = P(D = d)P(I = i)P(E = e)P(R = r|D = d, I = i)P(C = c|D = d, I = i, E = e).
 \end{aligned}$$

Para, por exemplo, determinarmos a probabilidade de saber o conceito C dado que acertou na resposta R , temos de efetuar os cálculos apresentados a seguir, utilizando a fatorização de probabilidades e a regra de Bayes (condicionada e a probabilidade total).

$$\begin{aligned}
 P(C = 1|R = 1) &= \frac{P(C = 1, R = 1)}{P(R = 1)} \\
 &= \frac{\sum_{d,i} P(C = 1, R = 1, D = d, I = i)}{\sum_d P(R = 1, D = d)} \\
 &= \frac{\sum_{d,i} P(D = d)P(I = i)P(R = 1|D = d, I = i)}{\sum_{d,i} P(D = d)P(I = i)P(R = 1|D = d, I = i)} \\
 &= 0.6829.
 \end{aligned}$$

O problema da especificação de parâmetros é um problema bem conhecido por pes-

quisadores em redes bayesianas. A dificuldade do processo de aquisição de conhecimento (determinação destes parâmetros) torna-se um maior inconveniente no uso de redes bayesianas, juntamente com a dificuldade de algoritmos de propagação. Apresentamos uma abordagem que permite que outras pessoas possam fornecer novos parâmetros com significado intuitivo, o que simplifica a especificação.

5.3.4 Simplificação de parâmetros a especificar para nós que representam evidências (provas)

Nesta secção descrevemos as relações entre nós conceito e nós evidência, em que estas evidências representam perguntas tipo teste. Mostramos, também, uma forma de simplificar o problema de especificação dos parâmetros.

A rede que modela as relações entre conceitos e perguntas tipo teste é a que se segue:

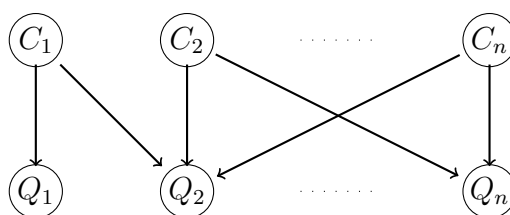


Figura 5.10: Relações entre conceito – pergunta tipo teste (Itens).

Como dito anteriormente, os parâmetros necessários para esta rede seriam o conjunto de probabilidades à priori para os nós conceitos (C_1, C_2, \dots, C_n) , onde necessitaríamos de n valores e a probabilidade condicionada do cada nó evidência (Q_i) dados os seus respetivos pais. Para simplificar na medida do possível estes parâmetros, vamos descrever uma proposta que se baseia no uso da função G usada em TRI (teoria de resposta ao item).

Propõe-se, então, uma abordagem baseada no cálculo das probabilidades usando uma função G definida para esta finalidade, em que, para cada pergunta item Q_i , quatro parâmetros $\{c, s, b, a\}$ têm de ser especificados pelo professor, sendo eles:

- c é o fator de adivinhar (em inglês *guess*), o que representa a probabilidade de que um aluno sem o conhecimento adivinhe a resposta correta ao item Q_i . Em questões

de escolha múltipla este valor é $\frac{1}{n}$ em que n é o numero de opções no item.

- s é o fator de descuido (em inglês *slip*), o que representa a probabilidade de que um aluno com todos os conhecimentos necessários não responde corretamente à pergunta item Q_i .
- b é o nível de dificuldade, que estima o quão difícil é o item de Q_i .
- a é o índice de discriminação, que representa o grau em que o item discrimina os alunos (o valor geralmente utilizado é 0.3).

A função G é então usada para calcular as probabilidades condicionadas. Esta abordagem reduz o número de parâmetros necessários de uma questão que envolve m conceitos, de 2^m para $4 + m$, os quatro acima mais o peso de cada conceito na questão. A função G fornece uma curva suave para atribuir essas probabilidades condicionadas, cobrindo a faixa entre c e $1 - s$.

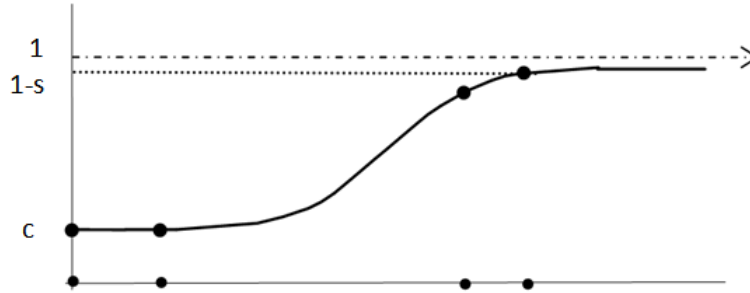


Figura 5.11: Gráfico da função $G(x)$ usada para calcular as probabilidades condicionadas.

Intuitivamente, como ilustrado na Figura 5.11 isto significa que a probabilidade de o aluno dar uma resposta correta assumindo que sabe todos os conceitos associados a ela, é um menos o valor do parâmetro de “deslizamento” e, a probabilidade de dar uma resposta correta sem saber nada sobre os conceitos relacionados, é o valor do parâmetro “adivinhar”. Em geral, dispomos

$$P(Q = 1 | \bigcap_{C \in \Omega} C = 1) = 1 - slip,$$

e

$$P(Q = 1 | \bigcap_{C \in \Omega} C = 0) = guess,$$

onde Ω é o conjunto dos conceitos associados com a questão Q .

Temos que definir também as outras probabilidades condicionadas, para os casos onde o aluno conhece alguns conceitos relacionados com a questão, mas não outros. Para calcular estas probabilidades usamos a função G da teoria de resposta ao item (TRI),

$$G(x) = 1 - \frac{(1 - c)(1 + e^{-1.7ab})}{1 + e^{1.7a(x-b)}}.$$

onde a é o valor de discriminação, b o nível de dificuldade e c o valor de adivinhar. Temos que $G(0) = c$ e $\lim_{x \rightarrow \infty} G(x) = 1$. Às probabilidades restantes foram atribuídos valores crescentes entre c e $1 - s$. Desta forma, a probabilidade de uma resposta correta aumenta com o número de conceitos conhecidos e com maior peso.

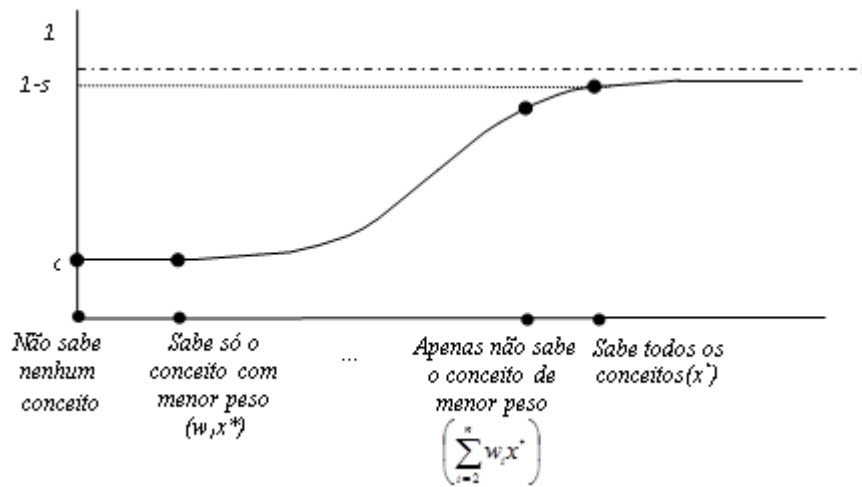


Figura 5.12: Uso do $G(x)$ para determinar as probabilidades.

A forma como usamos a função G é a seguinte: seja x^* tal que $G(x^*) = 1 - s$, donde vem que

$$x^* = \frac{\ln \left(\frac{(1-c)(1+e^{-1.7ab})-s}{s} \right) + 1.7ab}{1.7a}.$$

as probabilidades condicionadas são dadas de forma geral por

$$P\left(Q = 1 \mid \bigcap_{C \in \Omega_1} C = 1 \wedge \bigcap_{C \in \Omega_0} C = 0\right) = G\left(\left(\sum_{C \in \Omega_1} w_C\right) \times x^*\right),$$

onde $\Omega = \Omega_1 \cup \Omega_0$ é o conjunto de conceitos relacionados com a questão Q e w_C o peso que cada conceito tem na questão Q .

Se ordenarmos os conceitos envolvidos em cada questão por ordem crescente de importância (peso na questão), $w_1 \leq w_2 \leq \dots \leq w_n$, temos, por exemplo, $G(0) = c = \frac{1}{n}$ é a probabilidade de responder bem sem conhecer qualquer conceito, $G(w_1 x^*)$ é a probabilidade de responder corretamente conhecendo apenas o conceito menos importante (com menor peso na questão) e $G(\sum_{i=2}^n w_i x^*)$ é a probabilidade de responder bem dado que sabe todos os conceitos exceto o de menor peso.

Deste modo o professor só necessita de especificar o nível de dificuldade, o índice de discriminação e o peso de cada conceito na pergunta tipo teste, sendo que as probabilidades condicionadas são calculadas automaticamente com este método. Assim sendo, o método garante que na determinação das probabilidades condicionadas, quanto mais conhecimento mais provável é a escolha da resposta correta e, portanto, conduzirá a diagnósticos mais precisos.

Para mais detalhes sobre a função G , referimos a publicação original ([15]).

Apresentamos de seguida um exemplo que ilustra a funcionalidade do modelo definido.

Exemplo 5.5. [2]

Voltamos ao exemplo 5.3, cuja rede está representada na figura 5.13. As probabilidades condicionadas associados ao nó Q são calculadas usando os parâmetros “slip”, “acho”, “nível”, “discr” e os pesos dos conceitos D e I na questão Q , sendo estes 0.3 e 0.7, respetivamente, e com $s = 0.2$, $c = 0.25$, $a = 1.2$ e $b = 4$. Sendo, então, a função G para esta questão dada por:

$$G(x) = 1 - \frac{0.75(1 + \exp^{-8.16})}{1 + e^{2.04(x-4)}} = 1 - \frac{0.7502}{1 + e^{2.04(x-4)}}.$$

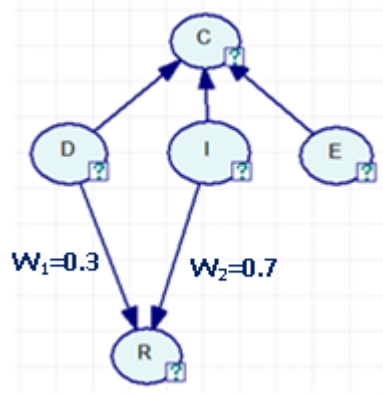


Figura 5.13: Rede bayesiana com evidência.

Obtemos, assim, que: $P(Q = 1|D = 1, I = 1) = 1 - s = 0.8$,
e $P(Q = 1|D = 0, I = 0) = c = 0.25$.

Como,

$$x^* = \frac{\ln \left(\frac{(1-0.25)(1+e^{-1.7 \times 1.2 \times 4})-0.2}{0.2} \right) + 1.7 \times 1.2 \times 4}{1.7 \times 1.2} \simeq 4.5,$$

obtemos as restantes probabilidades condicionadas

$$P(Q = 1|D = 1, I = 0) = G(w_1 x^*) = G(0.3 \times 4.5) = G(1.35) = 1 - \frac{0.7502}{1 + e^{2.04(1.35-4)}} \simeq 0.253,$$

$$P(Q = 1|D = 0, I = 1) = G(w_2 x^*) = G(0.7 \times 4.5) = G(3.15) = 1 - \frac{0.7502}{1 + e^{2.04(3.15-4)}} \simeq 0.362.$$

Usando esse modelo, todas as probabilidades condicionadas podem ser calculadas a partir deste pequeno conjunto de parâmetros. Tendo definido as probabilidades podemos calcular as crenças, após a obtenção de uma evidência. Isto é feito simplesmente por marginalizar, fatorizar e usar a regra de Bayes. Por exemplo,

$$\begin{aligned} P(D = 1|Q = 1) &= \frac{P(D = 1, Q = 1)}{P(Q = 1)} \\ &= \frac{\sum_i P(D = 1, R = 1, I = i)}{\sum_{d,j} P(R = 1, D = d, I = i)} \\ &= \frac{\sum_i P(D = 1)P(I = 1)P(R = 1|D = 1, I = i)}{\sum_{d,j} P(D = d)P(I = i)P(R = 1|D = d, I = i)}. \end{aligned}$$

Este valor pode ser calculado usando as probabilidades na rede bayesiana. Assim, a partir da evidência $Q = 1$, ou seja, o aluno ter dado uma resposta correta, uma nova crença ou probabilidade para o conceito D é obtida. A probabilidade $P(I = 1|Q = 1)$ é calculada de forma análoga e $P(E = 1|Q = 1) = P(E = 1)$. Também $P(C = 1|Q = 1)$ pode ser calculada e, em geral, depois da introdução de uma evidência na rede, as evidências são propagadas através de toda a árvore de conceitos.

Os algoritmos para fazer essa propagação são implementados no pacote *Smile* e nós usamo-lo na aplicação a descrever no próximo capítulo, onde explicamos como este modelo MAGRB foi implementado numa aplicação *web*, existente na UA, denominada SIACUA (Sistema Interativo de Aprendizagem da Universidade de Aveiro), que inclui um sistema autónomo de aprendizagem interativo.

Capítulo 6

A aplicação web SIACUA

A aprendizagem em matemática é considerada, por grande parte dos estudantes, como uma das mais difíceis. Muitos dos estudantes, quando chegam ao ensino superior, apresentam problemas de falta de pré-requisitos para os conteúdos exigidos, o que dificulta o entendimento até mesmo de conceitos mais básicos.

Em disciplinas de matemática, existentes na universidade de Aveiro, como, por exemplo, Cálculo I, Cálculo II e Cálculo III, existem alunos provenientes de vários cursos na área das ciências e engenharias, incluindo alunos que não frequentaram a disciplina de matemática A no ensino secundário, o que implica uma preparação menos boa. Existem, também, alunos que já não estudam há algum tempo e, por isso, os conceitos mais básicos estão esquecidos. Além disso, o tempo limitado para lecionar as disciplinas inibem o aprofundamento adequado dos assuntos e o grande número de alunos por turma, não permite que o professor disponha de tempo para um acompanhamento permanente e personalizado em função das dificuldades de cada aluno. Diante disto, cabe ao professor identificar disparidades de conhecimentos e adaptar os conteúdos pelo menos à maioria dos estudantes, o que nem sempre é possível dado que os programas estão definidos e têm de ser cumpridos, independentemente, do tipo de alunos, fazendo com que os que vem menos preparados se possam perder no processo de ensino- aprendizagem.

Como já referido no capítulo I, com os Sistemas Tutores Inteligentes (STIs), podemos

proporcionar um processo de aprendizagem individualizado em que cada aluno trabalha ao seu próprio ritmo, tendo disponíveis possíveis resoluções de exercícios, bem como a teoria necessária para a sua realização. Sendo que, baseado em redes bayesianas, o sistema permite ao aluno e ao professor diagnosticar os conhecimentos do aluno e a partir deste diagnóstico estabelecer uma sequência de passos eficazes para a sua aprendizagem.

Dada a grande capacidade de inferir probabilidades a partir de dados incertos e imprecisos, as redes bayesianas têm sido utilizadas para modelar o conhecimento do aluno em STIs. Neste capítulo apresentamos a implementação de uma RB, para um modelo do aluno, para mapear o seu conhecimento no domínio da disciplina de Cálculo III, sendo que o mesmo modelo, também já foi implementado para outras disciplinas, nomeadamente, Cálculo II, ensino secundário e ensino básico e está em permanente atualização.

Este modelo, denominado, SIACUA - Sistema Interativo de Aprendizagem por Computador, Universidade de Aveiro, à semelhança do modelo do aluno apresentado por Castillo et al. [4], utiliza como evidência, apenas, o resultado final da resposta dada pelo aluno a cada questão (correta ou incorreta) e não os passos que os alunos realizam até à sua solução, o que, por sua vez, seria mais vantajoso, pois assim o modelo identificaria com mais precisão o nível de conhecimento do aluno. Contudo, este trabalho diferencia-se por utilizar um modelo de RB mais genérico, não havendo necessidade de definir-se manualmente um grafo para cada tipo de problema. Porém, a diferença entre estes dois modelos reside no facto de que o SIACUA abrange um domínio geral (todos os temas de uma disciplina) e não um tema específico (equações do 1º grau) no caso do modelo de Castillo et al. [4].

Nesta secção, começaremos no tópico 6.1, por apresentar a aplicação referida, bem como os seus principais responsáveis. O Mapa Conceptual no domínio de Cálculo III, que auxiliou a construção da RB do MA proposto é descrito no tópico 6.2. A RB e as etapas envolvidas na sua modelação, nomeadamente os parâmetros a definir a sua propagação de conhecimento para os diferentes tipos de nós (conceito/ evidência) são apresentadas no tópico 6.3. Passaremos para o tópico 6.4, onde explicamos a inserção das questões assim como a sua proveniência, para tal abordaremos outros projetos implementados na universidade de Aveiro, nomeadamente PmatE e MEGUA. No tópico 6.5, faremos uma

breve síntese das aplicações informáticas (ferramentas) necessárias como pré-requisitos e a implementação do modelo. Finalmente uma breve discussão sobre os resultados das experiências já realizadas, bem como aspetos a melhorar, encontram-se no tópico 6.6.

6.1 Apresentação da aplicação

A aplicação foi desenhada e criada pelos docentes Luís Descalço (responsável pela programação) e Paula Carvalho (responsável pelos conteúdos), incluindo trabalho de diversos docentes e estudantes da Universidade de Aveiro e pode ser acedida em <http://siacua.web.ua.pt/>. Apresentamos a página inicial desta aplicação na figura 6.1.

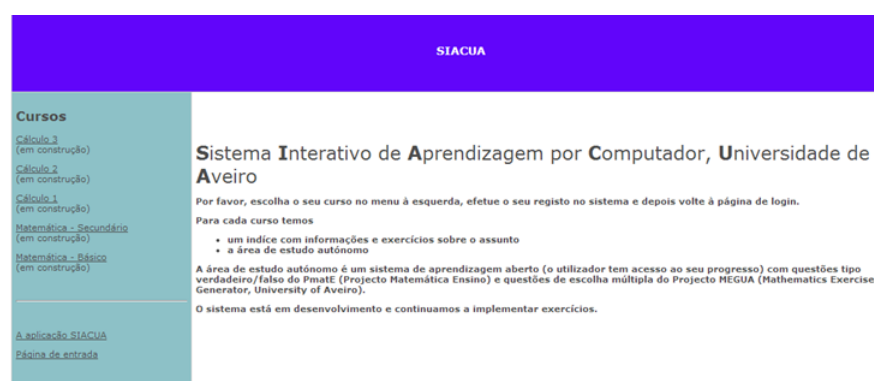


Figura 6.1: Página inicial da aplicação SIACUA.

A Programação do SIACUA e a inserção dos conteúdos de Modelos do PmatE em Cálculo III estão a cargo de Luís Descalço. Já as questões do MEGUA, são trabalho de Paula Carvalho, Luís Descalço, Dina Seabra, estando a programação destas a cargo de João Pedro Cruz.

Ao acedermos à disciplina Cálculo III, aparece-nos uma breve informação relativa às ferramentas necessárias para a aplicação correr com êxito, como ilustra a figura 6.2.

SIACUA é um sistema aberto, no sentido de que o aluno pode ver o seu progresso, o qual é mostrado na forma de barras de progresso como ilustrado na figura 6.3 [2].

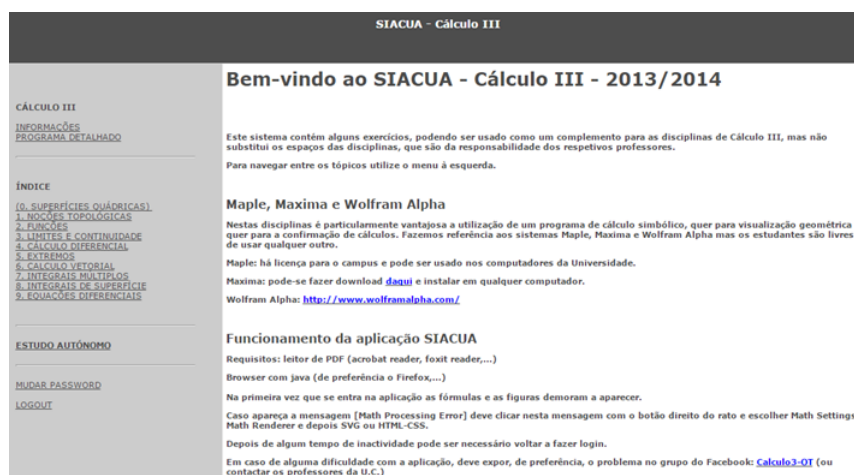


Figura 6.2: Página da SIACUA relativa ao Cálculo III.

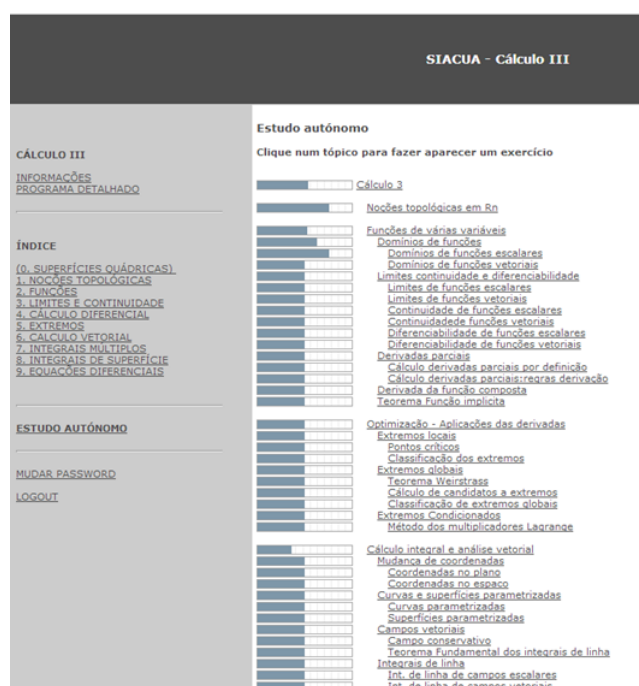


Figura 6.3: Progressos na SIACUA.

Cada barra no gráfico representa uma probabilidade: a “crença” na rede Bayesiana associada ao utilizador, correspondente ao respetivo tópico, com base nas evidências fornecidas. Intuitivamente, um valor p representado por uma barra significa apenas que, com as evidências até agora fornecidas, a probabilidade de o utilizador entender o tópico

correspondente é igual a p (por isso, antes de qualquer evidência, a probabilidade é 0,5)[23].

O aplicativo pode ser utilizado em computadores e dispositivos móveis. Os alunos usam-no como uma ferramenta para estudar e os professores usam-no, às vezes, nas aulas. Um professor pode ver o progresso de cada aluno na aplicação e também o progresso médio de cada turma em cada tema a ser estudado em [2].

A interação é mínima, consistindo em responder a perguntas de verdadeiro / falso do PmatE e perguntas de escolha múltipla do MEGUA, que descreveremos com mais pormenor no tópico 6.4 deste capítulo, podendo ser o aluno, além de selecionar o conceito, a selecionar as perguntas. Esta seleção de perguntas pode ser feita clicando na barra de progresso correspondente, o que de forma aleatória seleciona uma questão relacionada com o conceito correspondente, ou através da introdução de o número da questão, como ilustrado na figura 6.4 [2].

Figura 6.4: Escolha das questões na SIACUA.

Os números das questões só são mostrados ao aluno no momento em que ele está a responder à pergunta, o que lhe permite identificar a questão. Com a questão identificada ele pode seleccioná-la para responder novamente ou discuti-la, interagindo com os professores e colegas [2].

O objetivo é inferir o estado cognitivo do aluno, a partir dos dados coletados, diagnosticar o seu estado no processo de aprendizagem. O que por vezes não é tarefa trivial, uma vez que os dados são incompletos, pois há imensos fatores (variáveis) externos que não é possível ter em conta no diagnóstico, por exemplo, a ajuda de outra pessoa, as características do meio onde é acedida a aplicação, o estado do aluno, etc. Este é um passo difícil de implementar em sistemas de aprendizagem interativa. No entanto, este aplicativo fornece uma maneira muito simples de, apresentando questões ao aluno, ele obter algum *feedback*. O estudante sabe imediatamente se a sua resposta está certa ou errada, pode ver a solução

detalhada e tem informações sobre o seu progresso geral no curso, dado pelas barras de progresso, como ilustrado na figura 6.3 [2].

No índice detalhado estão os tópicos que constam do programa da unidade curricular Cálculo III. Nalguns casos o aluno poderá encontrar exercícios propostos e/ou resolvidos. Para a disciplina de cálculo III, os alunos além do sistema de estudo com exercícios tipo verdadeiro/falso (PmatE) e exercícios de escolha múltipla com resolução detalhada (ME-GUA), têm acesso a exercícios com as respetivas soluções e a um resumo sobre quadráticas. Terminamos este tópico com uma frase retirada de <http://siacua.web.ua.pt/calculo2/> e passaremos para a apresentação do mapa concetual da disciplina de calculo 3.

“Na secção estudo autónomo poderá responder a exercícios de escolha múltipla e ir analisando a evolução do seu conhecimento. Se a barrinha correspondente a um dado tópico aumentar isso significa que o seu conhecimento sobre esse conteúdo aumentou também. O desafio é o de os alunos conseguirem levar as barrinhas aos 100 % e assim divertirem-se a estudar!”

6.2 Mapa conceptual no domínio de Cálculo III

“Esta aplicação implementa um modelo do aluno baseado em RB, juntamente com um algoritmo para a propagação do conhecimento. Cada aluno tem uma rede Bayesiana associada, onde cada nó representa um tópico, embora às vezes chamam-lhe um nó conceito para distinguir dos nós de provas. Inicialmente, antes de qualquer interação do aluno, ou seja antes de começar a responder às perguntas, a sua rede Bayesiana contém apenas conceitos, que é o mapa conceptual do domínio a ser estudado” [2].

O primeiro passo para o desenvolvimento deste Modelo do aluno no domínio da disciplina de calculo III, foi a definição dos assuntos, dos temas envolvidos em cada assunto e por sua vez dos conceitos relacionados com cada tema. Esse mapa conceptual tem a estrutura de um grafo orientado acíclico, definido de modo a sistematizar as referidas relações dos conhecimentos. É portanto, um diagrama que permite visualizar as relações hierárquicas e dependentes entre os temas dado os conceitos e, por sua vez, os assuntos

dado os temas. Mostram, portanto, as relações de subordinação e superordenação entre os conceitos que possivelmente afetarão um tema e que por sua vez afetarão um dado assunto, o que é bastante útil para a construção da rede bayesiana. Dessa forma, a partir do levantamento realizado pelos professores do domínio (neste exemplo de calculo III), foi elaborado o mapa conceptual da disciplina. Este mapa representa os temas necessários para saber todos os conteúdos da disciplina, como ilustrado na figura 6.5.

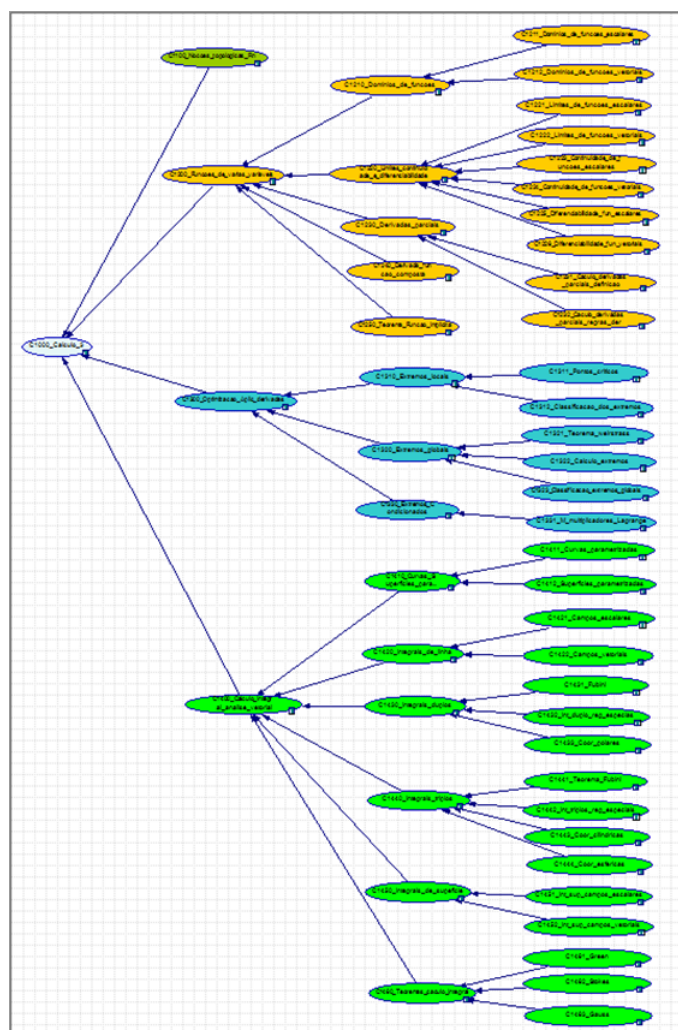


Figura 6.5: Mapa conceptual: rede bayesiana com nós que representam conceitos.

Este modelo é dividido em níveis, em que a variável topo é a que representa o domínio. Para todos estes níveis visíveis as variáveis são não observáveis, isto é, não podem ser

medidas diretamente. Serão inferidas pela RB, quando há inserção de evidências pelo aluno.

Cada vez que o aluno responde a uma pergunta, um novo nó é adicionado à rede, chamado nó de provas, sendo conectado com os conceitos que a pergunta envolve. Estes nós fornecem evidência de conhecimento sobre um determinado conceito. Ou seja, estas variáveis são aquelas que podem ser medidas diretamente e para as quais são obtidas evidências diretas e o conhecimento do estudante é propagado e atualizado, como veremos no tópico seguinte.

Os estados das variáveis, conforme demonstra a figura, foram definidos de modo que cada variável conceito possui dois estados, que representa o conhecimento do aluno sobre um conceito específico: Sabe ou Não sabe a que corresponde os valores 1 ou 0, respectivamente. A implementação foi feita com recurso ao GeNIe & SMILE.2.

Mais à frente descreveremos como esses estados são medidos e atualizados. Para saber o estado da variável domínio calculo III (C_{1000}), temos de saber os estado das variáveis assuntos C_{1100} , C_{1200} , C_{1300} , C_{1400} , que correspondem a Noções_Topológicas, Funções_de_Várias_Variáveis, Optimização_Aplic_Derivadas e Cálculo_Integral_Análise_Vectorial, respetivamente.

Para ilustrar, mostramos agora somente um excerto do mapa, sem perda de generalidade, escolhemos a modelação do tema C_{1300} _Optimização_Aplic_derivadas representada na figura 6.6.

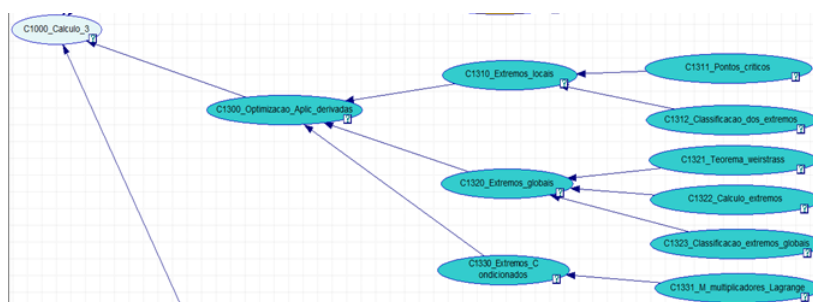


Figura 6.6: Excerto do mapa conceptual.

Podemos, então, constatar que para o aluno saber o assunto “Optimização_Aplicações

das derivadas (C_{1300})” precisa saber os temas “Extremos Locais (C_{1310})”, “Extremos Globais (C_{1320})” e “Extremos Condicionados (C_{1330})” e mostra também que, por exemplo para saber o tema “Extremos Locais (C_{1310})” precisa saber os conceitos “Pontos críticos (C_{1311})” e “Classificação de extremos (C_{1312})”.

O Mapa Concetual desenvolvido é a base para a criação da RB que compõe o Modelo do aluno no STI. A modelação da RB é detalhada no tópico seguinte.

6.3 Modelação da rede bayesiana

O primeiro passo na construção de uma RB é a definição das variáveis do domínio de conhecimento e suas relações. Posteriormente prossegue-se com a definição das probabilidades nas relações de causa-efeito entre as variáveis. Essas constituem provavelmente as principais fases no desenvolvimento de um STI baseado em redes bayesianas.

6.3.1 Na rede de conceitos

Analogamente ao estudado no capítulo 5 desta dissertação, para um MGARB (modelo genérico do aluno baseado em redes bayesianas), as probabilidades condicionadas de cada conceito, podem ser calculadas automaticamente a partir dos pesos de cada conceito (Pai) que lhe está associado.

Depois do mapa conceptual construído definiu-se os pesos para as relações de agregação, que tomaram os valores de $\frac{1}{k}$, sendo k o número de conceitos agregados. Ou seja, os pesos dos conceitos nos temas são igualmente repartidos, e assim sucessivamente até ao domínio. Para definirmos as probabilidades de cada nó, clicamos no canto inferior esquerdo de cada um dos nós e obtemos as respetivas janelas onde podemos definir as suas propriedades e definições. Inicialmente as probabilidades marginais de cada nó são definidas com 50% para cada um dos estados, sabe e não sabe, da variável, como ilustrado no excerto da rede representado pela figura 6.7.

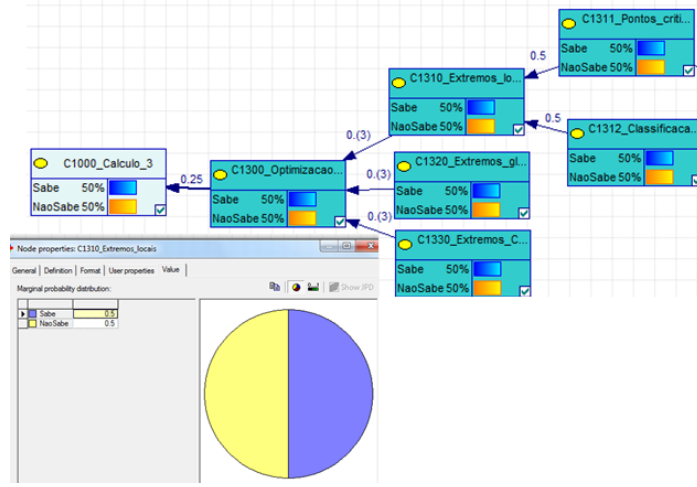


Figura 6.7: Excerto de uma rede com as probabilidades marginais.

Por exemplo, nesta figura, podemos ver, além dos nós que representam os temas, a janela com os valores das probabilidades marginais á priori, para o nó C1310_Extremos Locais.

Recapitulando, agora, o exposto no capítulo anterior, os pesos que medem a importância, relativa a cada conceito C_i , ao tema a que estão agregados, são definidos tal que $\sum_{i=1}^m w_i = 1$. As probabilidades condicionadas dos nós assuntos (filhos) dado os respectivos temas agregados (pais) é calculada da seguinte maneira:

$$P(T = 1 | \{C_i = 1, i \in I\}, \{C_i = 0, i \notin I\}) = \sum_{i \in I} w_i, \text{ onde } I = \{i \in \{1, 2, \dots, m\} \text{ tal que } C_i = 1\}.$$

De seguida apresentamos dois exemplos, do excerto anterior, que mostram como as probabilidades condicionadas à priori são definidas, antes de qualquer evidência dada pelo aluno.

Exemplo 6.1.

No tema C1310_Extremos_Locais que é composto pelos conceitos C_{1311} e C_{1312} os pesos são $w_{1311} = \frac{1}{2}$ e $w_{1312} = \frac{1}{2}$ respetivamente. Estas ponderações são utilizadas para definir a tabela de probabilidade condicionada do nó C_{1310} dado seus pais C_{1311} e C_{1312} , como explicado no capítulo anterior a tabela da figura 6.8 mostra as probabilidades condicionadas resultantes.

		C1311_Pontos...		C1312_Classifi...	
		Sabe	NaoSabe	Sabe	NaoSabe
► Sabe		1	0.5	0.5	0
NaoSabe		0	0.5	0.5	1

Figura 6.8: Tabela de probabilidades condicionadas do nó C_{1310} .

Exemplo 6.2.

No tema $C_{1320_Extremos_Globais}$ que é composto pelos conceitos C_{1321} , C_{1322} e C_{1323} os pesos são $w_{1321} = \frac{1}{3}$ e $w_{1322} = \frac{1}{3}$ e $w_{1323} = \frac{1}{3}$, respetivamente. Estas ponderações são utilizadas para definir a tabela de probabilidade condicionada do nó C_{1320} dado seus pais C_{1321} , C_{1322} e C_{1323} , como explicado no capítulo anterior a tabela apresentada na figura 6.9 mostra as probabilidades condicionadas resultantes.

		C1321_Teore...		C1322_Calcul...		C1323_Classifi...	
		Sabe	NaoSabe	Sabe	NaoSabe	Sabe	NaoSabe
► Sabe		1	0.66666667	0.66666667	0.33333333	0.66666667	0.33333333
NaoSabe		1.00000001	0.33333333	0.33333333	0.66666667	0.33333333	0.66666667

Figura 6.9: Tabela de probabilidades condicionadas do nó C_{1320} .

Os valores utilizados no preenchimento das tabelas de probabilidades associadas a cada um dos nós foram definidos uniformemente para a disciplina de cálculo III. Para as outras disciplinas já foram definidos de forma empírica, sendo, na maioria, pesos diferentes nos diferentes conceitos relacionados. Ou seja as tabelas de probabilidades para cada variável na RB, foram definidas com base nas crenças dos professores responsáveis sobre o conhecimento dos estudantes. A precisão das inferências realizadas pela rede deve ser encontrada a partir de testes e ajustes, pois não é possível definir métricas exatas para o comportamento humano.

Consideremos, agora, a figura 6.10, que representa um excerto da rede, com os respetivos pesos para cada conceito no tema a que estão agregados.

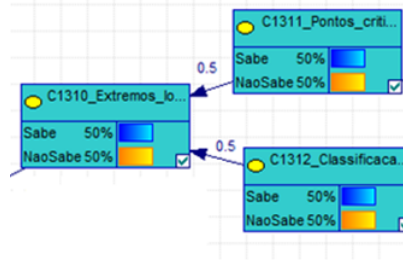


Figura 6.10: Excerto da rede com os respectivos pesos.

A distribuição conjunta das variáveis do modelo pode ser expressa via $P(C_{1310}, C_{1311}, C_{1312}) = P(C_{1311})P(C_{1312})P(C_{1310}|C_{1311}, C_{1312})$ lembrando que os conceitos C_{1311} e C_{1312} influenciam o conhecimento do tema C_{1310} .

Suponhamos que obtivemos a evidência de que o aluno sabe o conceito $C_{1311_Pontos_Criticos}$ e o tema $C_{1310_Extremos_Locais}$ e que se queria conhecer a probabilidade de saber o conceito $C_{1312_Classificação_de_Extremos}$, dadas as observações dos outros, esta probabilidade não está armazenada diretamente no modelo e é necessário calculá-la. O cálculo das probabilidades pode ser realizado mediante uma simplificação e fatorização de variáveis, como mostramos seguidamente:

$$P(C_{1312} = 1|C_{1310} = 1, C_{1311} = 1) = \frac{P(C_{1312} = 1, C_{1310} = 1, C_{1311} = 1)}{P(C_{1310} = 1, C_{1311} = 1)}.$$

Dadas as independências condicionadas a equação acima pode ser expressa como:

$$\begin{aligned} & \frac{P(C_{1311} = 1)P(C_{1312} = 1)P(C_{1310} = 1|C_{1311} = 1, C_{1312} = 1)}{P(C_{1310} = 1, C_{1311} = 1, C_{1312} = 1) + P(C_{1310} = 1, C_{1311} = 1, C_{1312} = 0)} \\ &= \frac{0.5 \times 0.5 \times 1}{0.5 \times 0.5 \times 1 + P(C_{1311} = 1)P(C_{1312} = 1)P(C_{1310} = 0|C_{1311} = 1, C_{1312} = 0)} \\ &= \frac{0.25}{0.25 + 0.5 \times 0.5 \times 0.5} \\ &= 0.(6). \end{aligned}$$

6.3.2 Na rede de conceitos-evidência

Os nós evidências representam as perguntas dadas pelo aluno e cada um será ligado aos conceitos da rede, necessários para o aluno responder corretamente. Como também já

estudado no capítulo anterior, para cada resposta seria preciso indicar as 2^n probabilidades condicionadas, em que n é o número de conceitos que a resposta envolve. Mas, de forma a simplificar estes parâmetros, estas probabilidades serão calculadas a partir de parâmetros: descuido - s (slip), adivinhar - c (guess), nível de dificuldade - b e índice de discriminação - a da função G , apresentada no MGARB, lembrando:

$$G(x) = 1 - \frac{(1 - c)(1 + e^{-1.7ab})}{1 + e^{1.7a(x-b)}}, \quad x \geq 0.$$

Uma vez que, nesta aplicação, as perguntas são do tipo escolha múltipla ou verdadeiro ou falso, temos que no primeiro caso o parâmetro “adivinhar - c ” é de 0.25, dado que existem quatro opções em que uma só está correta e no segundo caso será de 0.5. Ao parâmetro “descuido - s ” será atribuído o valor 0.2, o valor atribuído, pelo modelador, para o parâmetro “nível de dificuldade - b ” vai estar entre 0 e 5, para cada item e por fim para ao “índice de discriminação - a ” é atribuído, normalmente, o valor 0.3.

Apresentamos agora dois exemplos que ilustram como os valores das probabilidades condicionadas são calculados usando a função G .

Exemplo 6.3.

Seja Q uma pergunta do MEGUA que requer o conhecimento dos temas C1311_Pontos_Criticos e C1312_Classificação_dos_Extremos, com pesos $w_{1311} = 0.6$ e $w_{1312} = 0.4$ respetivamente, então um nó Q será adicionado à rede, como mostra a figura 6.11.

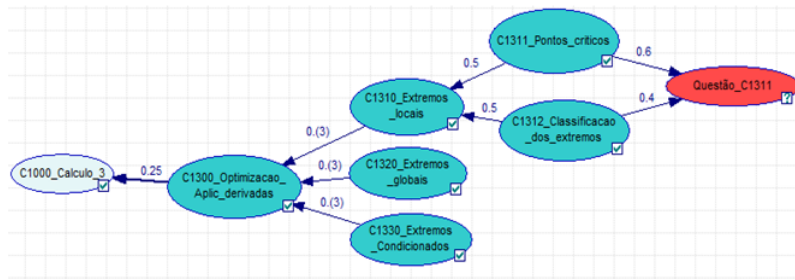


Figura 6.11: Excerto da rede com adição do nó Q do exemplo 6.3.

Uma vez que, nesta aplicação, as perguntas são do tipo escolha múltipla, $c = 0.25$ e os restantes parâmetros definidos são $s = 0.2$, $b = 1$ e $a = 0.3$. Sendo, então que, a função G

para esta questão, vem dada por:

$$G(x) = 1 - \frac{(1-c)(1+e^{-1.7ab})}{1+e^{1.7a(x-b)}} = 1 - \frac{1.2}{1+e^{0.51(x-1)}}.$$

Obtemos, assim, que:

$$P(Q = 1|C_{1311} = 1, C_{1312} = 1) = 1 - s = 0.8.$$

$$P(Q = 1|C_{311} = 0, C_{312} = 0) = c = 0.25.$$

$$x^* = \frac{\ln\left(\frac{(1-0.25)(1+e^{-1.7 \times 0.3 \times 1})-0.2}{0.2}\right) + 1.7 \times 0.3 \times 1}{1.7 \times 0.3} \simeq 4.156.$$

$$\begin{aligned} P(Q = 1|C_{1311} = 0, C_{1312} = 1) &= G(w_{1312} \times x^*) = G(0.4 \times 4.156) \\ &= G(1.6624) = 1 - \frac{1.2}{1+e^{0.51(1.6624-1)}} \\ &\simeq 0.5004. \end{aligned}$$

$$\begin{aligned} P(Q = 1|C_{1311} = 1, C_{1312} = 0) &= G(w_{1311} \times x^*) = G(0.6 \times 4.156) \\ &= G(2.49) = 1 - \frac{1.2}{1+e^{0.51(2.49-1)}} \\ &\simeq 0.618. \end{aligned}$$

A tabela 6.1 resume os valores das probabilidades condicionadas para o nó da questão, obtidos analiticamente e a tabela apresentada na figura 6.12 traduz as probabilidades calculadas pelo modelo.

C_{1311}	1		0	
C_{1312}	1	0	1	0
$P(Q = 1)$	0.8	0.618	0.5004	0.25

Tabela 6.1: Tabela de probabilidades condicionadas do nó Q , obtidas analiticamente.

Como podemos verificar, os valores são ligeiramente diferentes, devido ao facto de que na programação do modelo foi utilizada uma discretização da função G , para encontrar o valor de x^* .

C1311_Pontos...	Sabe		NaoSabe	
C1312_Classifi...	Sabe	NaoSabe	Sabe	NaoSabe
▶ Acertar	0.8038053	0.63007309	0.49332352	0.25
Erar	0.1961947	0.36992691	0.50667648	0.75

Figura 6.12: Tabela de probabilidades condicionadas do nó Q , obtidas no modelo.

Exemplo 6.4.

Seja Q uma pergunta do MEGUA que requer o conhecimento dos temas C_{1321} , C_{1322} e C_{1323} com pesos $w_{1321} = 0.1$, $w_{1322} = 0.6$ e $w_{1323} = 0.3$, respectivamente, então um nó Q será adicionado à rede, como mostra a figura 6.13.

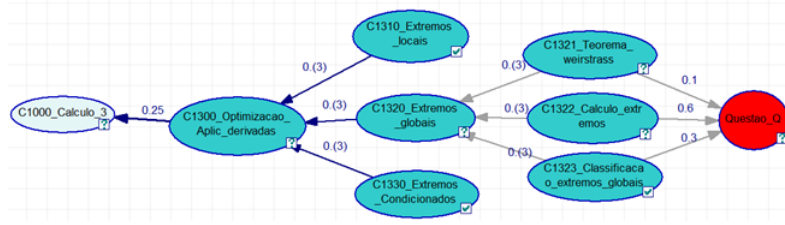


Figura 6.13: Excerto da rede com adição do nó Q do exemplo 6.4.

Uma vez que, nesta aplicação, as perguntas são do tipo escolha múltipla, os parâmetros a definidos são, a título de exemplo: $c = 0.25$, $s = 0.2$, $b = 3$ e $a = 1.2$.

Sendo a função G , para esta questão, dada por:

$$G(x) = 1 - \frac{0.75(1 + e^{-6.12})}{1 + e^{2.04(x-3)}} = 1 - \frac{0.752}{1 + e^{2.04(x-3)}},$$

e

$$x^* = \frac{\ln\left(\frac{(1-0.25)(1+e^{-1.7 \times 1.2 \times 3})-0.2}{0.2}\right) + 1.7 \times 1.2 \times 3}{1.7 \times 1.2} \simeq 3.49735.$$

A tabela de probabilidades condicionadas resultante para o item de teste Q é mostrada na tabela 6.2.

Geometricamente podemos constatar que $G(0) = c = 0.25$ e $\lim_{x \rightarrow \infty} G(x) = 1$. Às probabilidades restantes foram atribuídos valores crescentes entre c e $1 - s$. Desta forma, a probabilidade de uma resposta correta aumenta com o número de conceitos conhecidos, como ilustra a figura 6.14.

Conceitos que a questão envolve			Soma dos pesos dos conceitos que sabe	Valores de x	$P(Q = 1 C_{1321}, C_{1322}, C_{1323})$
C_{1321}	C_{1322}	C_{1323}	$\sum_{i \in \Omega_1} w_i$	$(\sum_{i \in \Omega_1} w_i) \times x^*$	$G(x)$
0	0	0	0	0.000	0.250
1	0	0	0.1	0.350	0.252
0	0	1	0.3	1.049	0.262
1	0	1	0.4	1.399	0.276
0	1	0	0.6	2.098	0.351
1	1	0	0.7	2.448	0.432
0	1	1	0.9	3.148	0.680
1	1	1	1	3.497	0.800

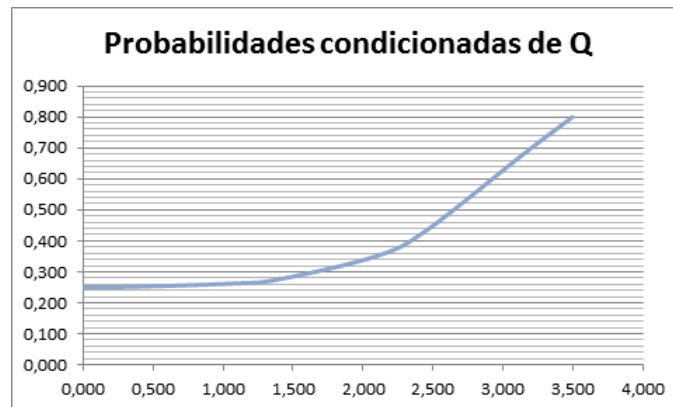
Tabela 6.2: Cálculo das probabilidades condicionadas usando a função G .

Figura 6.14: Probabilidades condicionadas de Q.

Em geral, pensa-se que, esta abordagem é fácil de implementar e fornece automaticamente todas as probabilidades condicionadas necessárias, que são calculadas com uma função dos quatro parâmetros estimados pelos professores [4].

Uma vez que um aluno responde a uma questão, a evidência sobre ela é usada para calcular as probabilidades posteriores. Como apresentado no capítulo anterior, este mecanismo de atualização é aplicação direta da lei de probabilidade total e da probabilidade condicionada, como se mostra no exemplo que se segue, e em que apresentamos uma comparação dos valores obtidos na aplicação com os calculados de forma analítica.

Exemplo 6.5.

Voltando ao exemplo 6.4 em que temos um excerto da rede bayesiana, para o tema C_{1300} , mostrado na figura 6.15 e em que mostramos o estado da BN antes do aluno ter respondido a uma questão associada aos conceitos C_{1311} e C_{1312} . A evidência obtida, respondendo à questão Q , além de atualizar o conhecimento destes conceitos, vai ser propagada para o tema, assunto e domínio relacionados. Neste caso irá atualizar o estado de conhecimento dos nós C_{1310} , C_{1300} e C_{1000} :

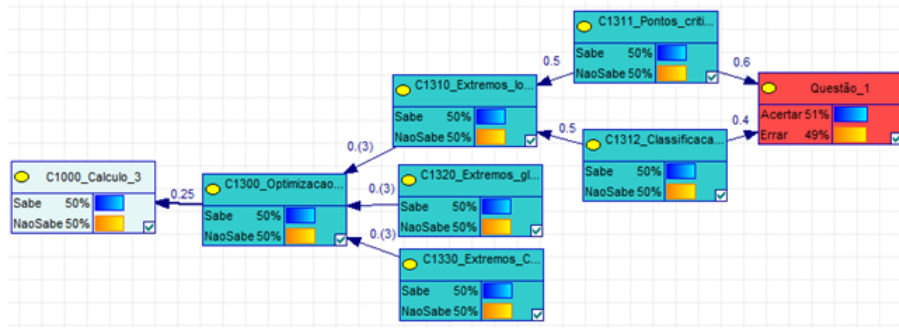


Figura 6.15: Atualização do estado de conhecimento dos nós C_{1310} , C_{1300} e C_{1000} .

As probabilidades, à priori foram definidas anteriormente, como sistematizado na figura 6.16, que mostra todas essas probabilidades nos nós envolvidos (ao excerto anterior foram suprimidos os nós C_{1320} e C_{1330} , porque não vão sofrer alterações com esta evidência). Suponhamos que o aluno respondeu corretamente à questão, as novas probabilidades para os nós C_{1311} e C_{1312} , são calculadas da seguinte maneira:

$$\begin{aligned}
 P(C_{1311} = 1|Q = 1) &= \frac{\sum_i P(Q = 1|C_{1311} = 1, C_{1312} = i)P(C_{1311} = 1)P(C_{1312} = i)}{\sum_{i,d} P(Q = 1|C_{1311} = d, C_{1312} = i)P(C_{1311} = d)P(C_{1312} = i)} \\
 &= \frac{0.8 \cdot 0.5 \cdot 0.5 + 0.618 \cdot 0.5 \cdot 0.5}{0.8 \cdot 0.5 \cdot 0.5 + 0.618 \cdot 0.5 \cdot 0.5 + 0.5004 \cdot 0.5 \cdot 0.5 + 0.25 \cdot 0.5 \cdot 0.5} \\
 &= 0.654.
 \end{aligned}$$

$$\begin{aligned}
 P(C_{1312} = 1|Q = 1) &= \frac{\sum_i P(Q = 1|C_{1311} = i, C_{1312} = 1)P(C_{1311} = i)P(C_{1312} = 1)}{\sum_{i,d} P(Q = 1|C_{1311} = d, C_{1312} = i)P(C_{1311} = d)P(C_{1312} = i)} \\
 &= \frac{0.8 \cdot 0.5 \cdot 0.5 + 0.5004 \cdot 0.5 \cdot 0.5}{0.8 \cdot 0.5 \cdot 0.5 + 0.618 \cdot 0.5 \cdot 0.5 + 0.5004 \cdot 0.5 \cdot 0.5 + 0.25 \cdot 0.5 \cdot 0.5} \\
 &= 0.6.
 \end{aligned}$$

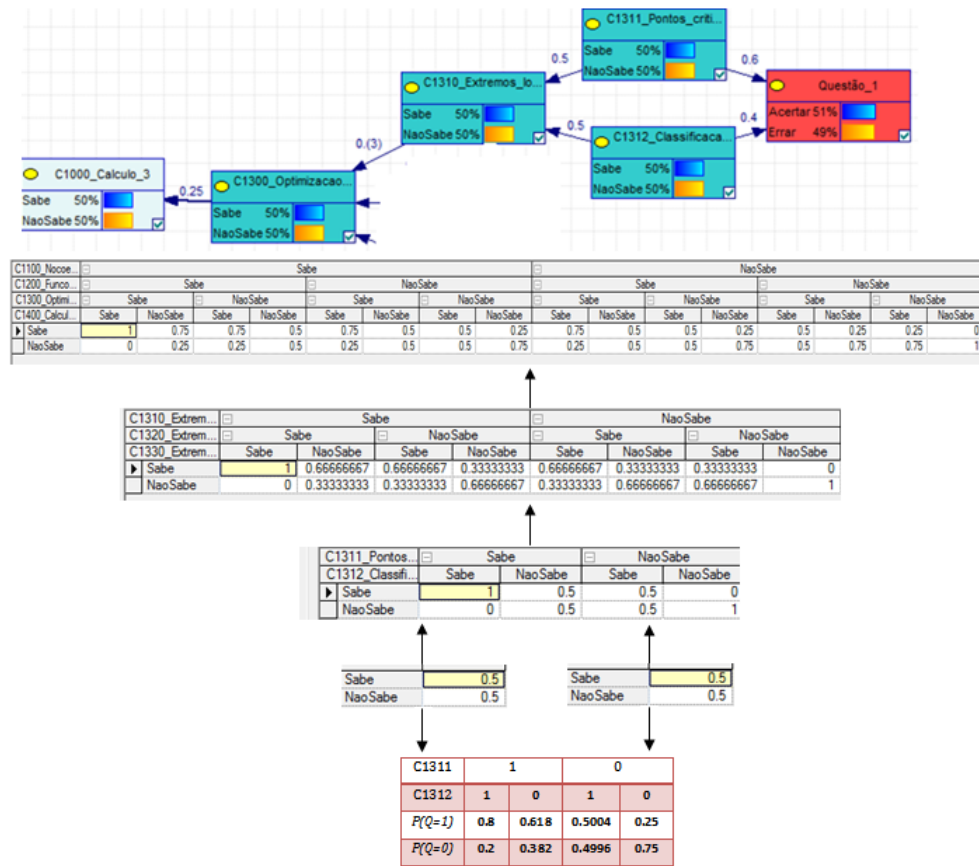


Figura 6.16: Probabilidades dos nós do exemplo 6.5.

Resumidamente as probabilidades marginais dos nós relativos aos conceitos C_{1311} e C_{1312} , encontram-se nas tabelas seguintes:

$P(C_{1311})$		$P(C_{1312})$	
=1	0.654	=1	0.6
=0	0.346	=0	0.4

Tabela 6.3: Probabilidades marginais dos nós C_{1311} e C_{1312} obtidas analiticamente.

A figura 6.17, mostra as probabilidades obtidas no GeNIe, que são ligeiramente diferentes das calculadas analiticamente, pela mesma razão referida atrás.

	Value
Sabe	0.65858769
NaoSabe	0.34141231

	Value
Sabe	0.59577792
NaoSabe	0.40422208

Figura 6.17: Probabilidades obtidas no GeNIe para C_{1311} e C_{1312} .

E assim são atualizadas as probabilidades dos nós relativos aos conceitos que a questão envolve. Para atualizar o nó relativo ao tema C1310_Extremos_Locais que envolve estes dois conceitos procede-se da seguinte forma:

$$\begin{aligned}
 P(C_{1310} = 1 | Q = 1) &= \frac{P(C_{1310} = 1, Q = 1)}{P(Q = 1)} \\
 &= \frac{\sum_{i,d} P(C_{1310} = 1, C_{1312} = i, C_{1311} = d, Q = 1)}{\sum_{i,d} P(Q = 1 | C_{1311} = d, C_{1312} = i) P(C_{1311} = d) P(C_{1312} = i)} = *.
 \end{aligned}$$

Pela teorema de fatorização sabemos que

$$P(C_{1310}, C_{1311}, C_{1312}, Q) = P(C_{1310} | C_{1311}, C_{1312}) P(C_{1311}) P(C_{1312}) P(Q | C_{1311}, C_{1312}),$$

donde vem

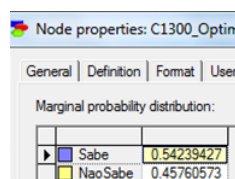
$$\begin{aligned}
 * &= \frac{\sum_{i,d} P(C_{1310} = 1 | C_{1311} = i, C_{1312} = d) P(C_{1311} = i) P(C_{1312} = d) P(Q = 1 | C_{1311} = i, C_{1312} = d)}{\sum_{i,d} P(Q = 1 | C_{1311} = d, C_{1312} = i) P(C_{1311} = d) P(C_{1312} = i)} \\
 &= \frac{1 \cdot 0.5 \cdot 0.5 \cdot 0.8 + 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.618 + 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5004 + 0 \cdot 0.5 \cdot 0.5 \cdot 0.25}{0.8 \cdot 0.5 \cdot 0.5 + 0.618 \cdot 0.5 \cdot 0.5 + 0.5004 \cdot 0.5 \cdot 0.5 + 0.25 \cdot 0.5 \cdot 0.5} \\
 &= \frac{0.3398}{0.5421} = 0.627.
 \end{aligned}$$

Comparando com o valor obtido no GeNIe, na figura 6.18 , podemos verificar que, arredondando às centésimas, não existe diferença.

	Value
Sabe	0.62718281
NaoSabe	0.37281719

Figura 6.18: Probabilidades obtidas no GeNIe para C_{1310} .

Tendo em conta as probabilidades condicionadas definidas à priori, da figura 6.16, a propagação da evidência para o nó tema C_{1300} será efetuada de modo análogo tendo-se obtido $P(C_{1300} = 1|Q = 1) = 0.542$. Observamos que a atualização das probabilidades do nó assunto, C_{1300} é aproximadamente igual à obtida no GeNIe, como podemos observar na figura 6.19.



Marginal probability distribution:	
Sabe	0.54239427
NaoSabe	0.45760573

Figura 6.19: Probabilidades obtidas no GeNIe para C_{1300} .

O aluno, depois desta evidência, ou seja depois de responder e ter acertado a esta única questão, já observa a seguinte alteração nas barras de progresso, da figura 6.20.

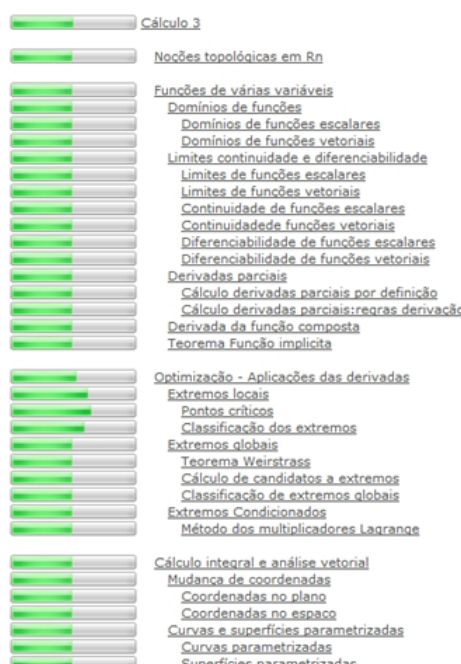


Figura 6.20: Barras de progresso do aluno.

Neste tipo de RB os nós evidência podem ser usados mais do que uma vez, ou seja, têm a capacidade de atualizar uma evidência, o que se vai propagar para toda a árvore de con-

ceitos e então todas as probabilidades na rede são atualizadas. Assim, o aluno, verificando o estado das barras de progresso pode fazer as suas inferências sobre o estado do seu conhecimento na disciplina e, assim, agir em conformidade. Por exemplo, poderá pedir ajuda ao professor e estudar melhor os respetivos conteúdos, de forma a colmatar as dificuldades nos assuntos, temas e conceitos que apresentam menos probabilidade de conhecimento (barras de progresso mais pequenas). Passamos agora para uma breve descrição da proveniência das questões do modelo.

6.4 Questões

A elaboração das questões, relativas aos vários assuntos em estudo, para inserir nestes tipos de modelo dos STIs, são normalmente uma limitação para a sua criação, devido ao grande número que é necessário para o modelo ser viável. Nesta aplicação isso não foi obstáculo dado que foram utilizadas questões já implementadas noutros projetos da universidade de Aveiro, nomeadamente PmatE [20] e MEGUA [13], para os quais faremos uma breve apresentação neste tópico, bem como o trabalho de estudantes de mestrado na criação de novas questões, tanto para Cálculo III e Cálculo II como para o nível básico e secundário.

6.4.1 PmatE

Apresentaremos uma breve descrição, deste projeto, baseada em [20] e em [2], para mais detalhe, como por exemplo, para consultar todos os seus projetos e iniciativas, podemos consultar a respetiva página representada pela figura 6.21 em <http://pmate4.ua.pt/hexeris/>.



Figura 6.21: Página do PmatE.

O PMatE (projeto de matemática ensino) foi criado na universidade de Aveiro, inicialmente com o principal objetivo de incentivar o gosto e sucesso em matemática e que devido ao envolvimento e cooperação de diversas entidades de ensino a nível nacional tomou dimensões internacionais através de projetos de cooperação com países de língua oficial portuguesa, alargando assim os seus objetivos, acompanhando a evolução tecnológica. Este projeto engloba agora diversas áreas de ensino, tornando-se num ousado e criativo projeto interdisciplinar, em que a sua principal função é a produção de objetos de aprendizagem, modelos geradores de questões e conteúdos digitais multimédia, disponibilizados numa plataforma de ensino assistido por computador, atualmente disponível apenas na Internet, abrangendo os vários graus de ensino, do básico ao superior.

“O PmatE é um projeto de Investigação e Desenvolvimento, que surgiu em 1989, com a preocupação de desenvolver ferramentas informáticas e conteúdos em diversas áreas do saber. A comemorar 25 anos, mantém como missão a aplicação de tecnologias e o desenvolvimento de conteúdos e eventos ao serviço da promoção do sucesso escolar e da cultura científica” (retirado de <http://pmate4.ua.pt/hexeris/index.php/home/quem-somos>).

Atualmente, neste projeto, existem 1252 perguntas parametrizadas disponíveis para a área de matemática e algumas são usadas na aplicação SIACUA. Estas são do tipo verdadeiro/ falso, agrupadas em lotes de quatro afirmações em que para cada um destes lotes consta um texto inicial. As questões são escolhidas aleatoriamente de um modelo, que se designa por Modelo gerador de questão - MGQ. Assim o estudante fornece quatro evidências na rede bayesiana.

As perguntas são parametrizados, o que significa que existem parâmetros na questão que são instanciados em tempo de execução pelo computador.

A figura 6.22 é um exemplo de uma questão gerada a partir de um MGQ. Neste exemplo, o centro de elipses e os seus semi-eixos são escolhidos aleatoriamente a partir de um conjunto de possibilidades especificadas na MGQ. Os parâmetros são utilizados para gerar o texto inicial e declarações, mas também as imagens, o que torna possível obter um rico conjunto de diferentes perguntas a partir de um único MGQ, usando a tecnologia do PmatE.

Seja C a curva fechada orientada positivamente, fronteira da região $D = \left\{ (x, y) \in \mathbb{R}^2 : \frac{1}{16}(x+4)^2 + \frac{1}{16}(y+4)^2 \leq 1 \right\}$. Considere o campo

$$F(x, y) = \left(\frac{1}{64}(x+4)^2(y+4), -\frac{5}{64}(x+4)(y+4)^2 \right).$$

Seja $f: D \rightarrow \mathbb{R}^2$ definida por $\left| -\frac{5}{64}(y+4)^2 - \frac{1}{64}(x+4)^2 \right|$. Então o volume do sólido ☐ V ☐ F ☐ NR

$V = \left\{ (x, y, z) \in \mathbb{R}^3 : (x, y) \in D \wedge 0 \leq z \leq f(x, y) \right\}$ é dado por $\left| \int F(x, y) \cdot d\vec{r} \right|$

A região D é conexa ☐ V ☐ F ☐ NR

O campo vectorial F não é conservativo. ☐ V ☐ F ☐ NR

$\int F(x, y) \cdot d\vec{r} = \int_D -\frac{5}{64}(y+4)^2 - \frac{1}{64}(x+4)^2 dx dy$ ☐ V ☐ F ☐ NR

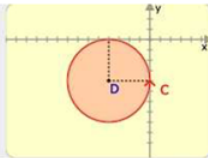


Figura 6.22: Uma questão de PmatE.

6.4.2 MEGUA

Apresentaremos uma breve descrição, deste projeto, baseada em [13] e em [2], para mais detalhe, podemos consultar a respetiva página da figura 6.23 em <http://cms.ua.pt/megua>.



Figura 6.23: Página do MEGUA.

O projeto MEGUA (Matemática Gerador de Exercício, Universidade de Aveiro) [13] foi criado na universidade de Aveiro desde 2012 e tem como função a criação de questões, maioritariamente em linguagem latex, em que a linguagem de programação, para a criação de valores, usada é Python com acesso às bibliotecas do ambiente gráfico de matemática Sage math, permitindo criar questões de escolha múltipla. Funciona como biblioteca onde são arquivados os diferentes “livros”(ficheiro de base de dados) que contêm os exercícios criados sobre os diversos assuntos de matemática. Permite que o enunciado de um exercício seja parametrizado, isto é, depois da sua produção permite através do identificador do

exercício, substituir os parâmetros e obter um novo exercício nos mesmos moldes, mas com valores ou funções diferentes. Isto é uma das mais-valias do modelo, tanto para o aluno, pois assim fica claro o que é essencial na questão, não dependendo dos valores dos parâmetros, como para o professor que pode pensar num modelo geral para a questão em vez de reutilizar as perguntas de exames anteriores fazendo pequenas alterações, o que é uma tarefa repetitiva e demorada.

O utilizador além de visualizar o enunciado, pode consultar a sua resolução, repetir o mesmo ou escolher outro aleatoriamente, tendo à sua disposição um índice de tópicos para escolha.

Com o Sage Notebook qualquer pessoa pode criar, colaborar e publicar conteúdos e exercícios interativos e enviá-los imediatamente para o aplicativo SIACUA.

Apresentamos de seguida um exemplo de uma questão do MEGUA na figura 6.24 bem como uma possível resolução caso o aluno necessite de a consultar.

6.4.3 Dinâmica dos três sistemas - MEGUA, PmatE e SIACUA

Numa primeira fase, foram seleccionadas, do grande número existente no PmatE, as perguntas que irão estar disponíveis para usar na aplicação Siacua. Essas, mesmas perguntas foram classificadas usando o mapa concetual, no sentido de atribuição de pesos para cada conceito que a questão envolve, bem como os parâmetros, c , a , b , e s . O acesso às respetivas questões do PmatE, através do SIACUA, é processado em tempo de execução, da seguinte forma: a aplicação SIACUA envia o pedido ao servidor *web* do PmatE, enviando o ID (identificador) da questão, este servidor gera a questão, usando o MGQ (Modelo gerador de questões), sendo que os parâmetros são gerados aleatoriamente no momento do pedido [2].

Toda a informação necessária para gerar as perguntas, os parâmetros e o estado atual da rede referente a cada aluno são armazenados numa base de dados.

Quanto às perguntas provenientes de MEGUA, estas são previamente gerados, ou seja, é criado e guardado na base de dados do SIACUA um conjunto de instâncias de cada

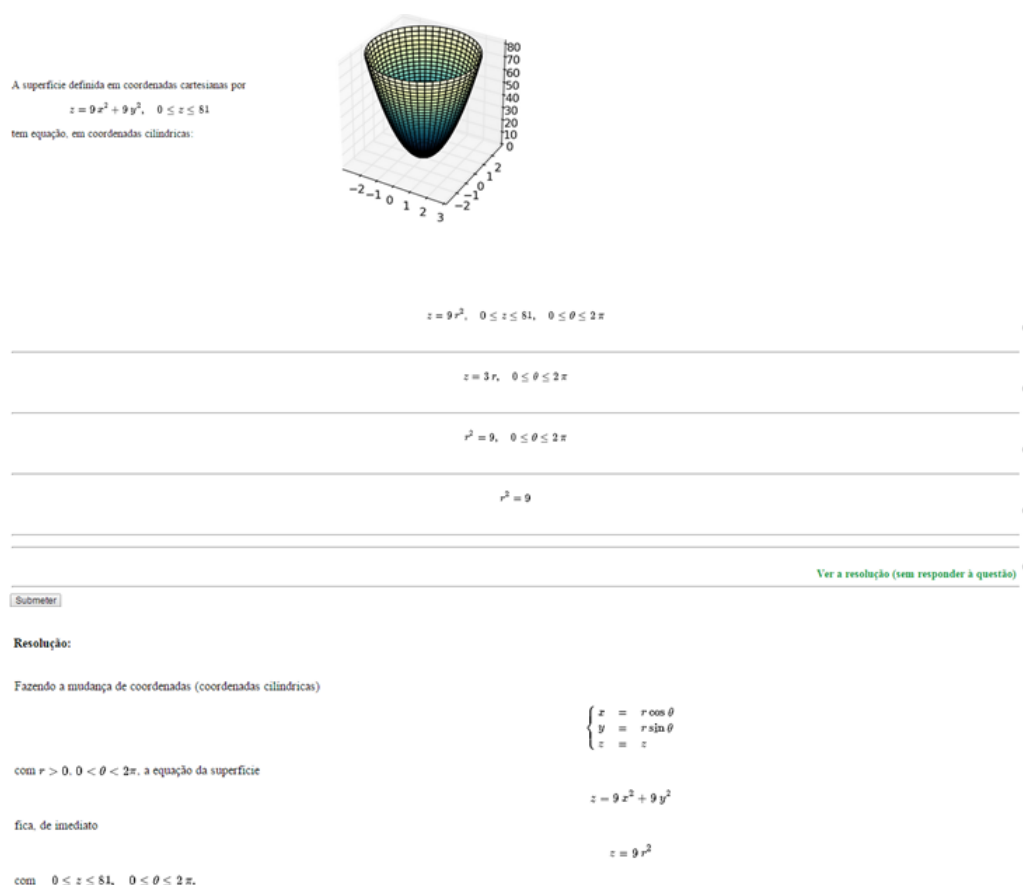


Figura 6.24: Uma questão de MEGUA.

pergunta. Esta forma de processar as questões tem vantagens e desvantagens. A principal desvantagem é que o número instâncias é fixo. Para ultrapassar esta desvantagem poder-se-ia definir um número elevado, mas assim ter-se-ia, custos ao nível de espaço na base de dados. Na prática, o resultado do uso deste processo é igual ao das questões geradas em tempo de execução, do PmatE.

Uma vantagem do processo usado para utilizar questões do MEGUA é a velocidade, já que as perguntas instanciadas já estão disponíveis. Outra vantagem é que o professor pode escolher as instâncias com melhores parâmetros, produzindo questões com melhor aparência e mais interessantes do que as obtidas, gerando os parâmetros aleatoriamente.

As questões são criadas dentro do Sage Notebook, usando o MEGUA, onde também

são selecionadas as instâncias com os parâmetros convenientes. De seguida associa-se a pergunta aos respetivos conceitos do mapa concetual do curso e definem-se os parâmetros da rede bayesiana, já estudados anteriormente, nomeadamente, s-descuido, c-adivinhar, a-discriminação e b- nível de dificuldade. Finalmente, a questão, as instancias e os parâmetros são enviados juntamente para o Siacua, ficando imediatamente disponível online.

6.5 Implementação de Software

Este modelo do aluno, baseado em redes bayesianas foi implementado com o pacote de Software GeNIe and SMILE, *freeware*. O GeNIe fornece um interface gráfica que nos permite desenhar a rede, de forma simples e intuitiva. O Smile permite utilizar redes bayesianas na programação e criar modelos probabilísticos e métodos de decisão sob a incerteza, para inclusão em STIs. Ambos os módulos, desenvolvidos no Laboratório de Sistemas decisão, da Universidade de Pittsburgh, foram colocados à disposição da comunidade em julho de 1998 e tem agora milhares de usuários em todo o mundo (<https://dslpitt.org/>).

A aplicação web SIACUA foi desenvolvida no Microsoft Visual Studio 2012 em C# com a base de dados em SQL.

Os professores desenharam o mapa conceitual, no ambiente de criação de questões do MEGUA, utilizaram HTML com Latex e Python para elaborar e disponibilizar as questões.

6.6 Experiências/ Resultados

Anteriormente, o mesmo modelo do aluno, em outras aplicações, foi proposto e testado com alunos simulados e posteriormente testado com alunos reais. Estes estudos poderão ser consultados em [4], [15] e [17], respetivamente.

A aplicação SIACUA é uma ferramenta complementar no processo de ensino-aprendizagem, não substituindo o ensino tradicional. Numa primeira avaliação do modelo foram analisados os dados obtidos na aplicação, como o número de *logins*, o número de exercícios acedidos e foi realizado um inquérito online sobre a aplicação. Podemos verificar

na figura 6.25 os resultados deste inquérito, bem como as questões em causa.

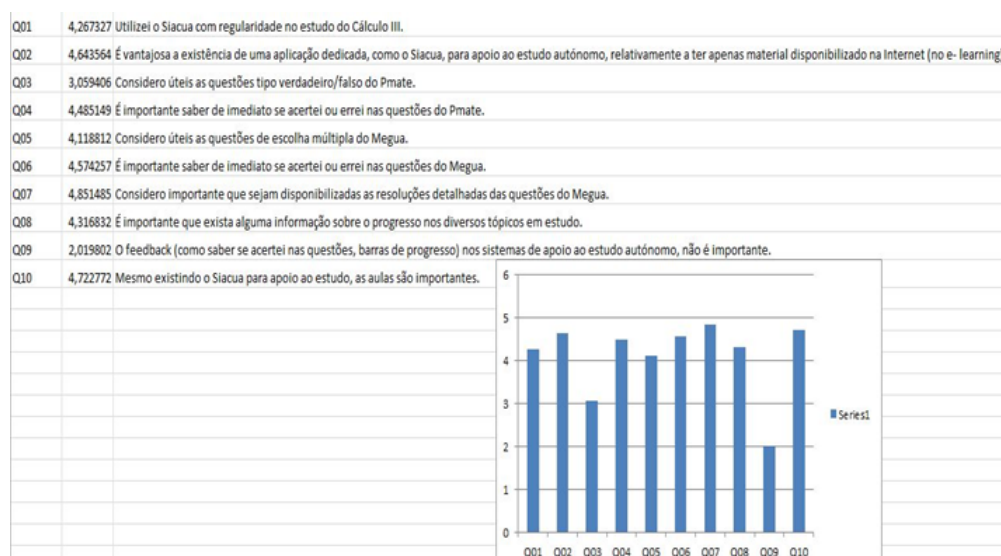


Figura 6.25: Resultados do inquérito.

Como podemos verificar, a maioria dos alunos considera muito importante a disponibilização da resolução detalhada das questões do MEGUA, talvez por esta razão a utilidade destas em comparação com as do PmatE, tenham tido uma maior percentagem de preferência pelos alunos.

Surgiu a necessidade de avaliar o modelo proposto comparando a aprendizagem do estudante mostrado na aplicação com a avaliação final na disciplina. Contudo, este estudo, só está a ser posto em prática, este ano letivo, porque nos anos anteriores não era viável, dado que não existia obrigatoriedade na utilização da aplicação. Este ano foi introduzida como parte da avaliação contínua, em que os alunos terão de responder a algumas questões sobre um tema que vai estar em avaliação em parte do semestre. Para tal, o professor, dá um tempo limite para a sua realização, em que nesse período de tempo o aluno não tem acesso à resolução das questões relativas aos conteúdos que estão em avaliação, ou seja o professor ocultou tanto a resolução como a possibilidade de escolher outra questão sem responder à selecionada. Mesmo que o aluno saia da aplicação, quando volta a última questão que não respondeu aparece-lhe novamente. Findo o prazo o aluno ficará com a nota correspondente à percentagem apresentada na barra de progresso do aluno, indicando-lhe

classificação que obteve.

Pretende-se depois verificar se existe correlação entre as variáveis “nota correspondente à barra de progresso no domínio da aplicação” com a “nota obtida no exame final”.

Pretende-se, assim, através de experiências, validar o modelo e definir possíveis adaptações que ele deva sofrer para aumentar o seu impacto positivo no processo de ensino-aprendizagem.

Contudo, os responsáveis pela aplicação, têm em mente algumas melhorias na aplicação, nomeadamente, estabelecer relações de pré-requisito, ou seja, quando o aluno não sabe um conceito por falta de bases lecionadas em disciplinas de matemática de anos anteriores, o sistema permitir saltar de cursos para outros e sugerir textos de complemento, guiar automaticamente o utilizador, o que seria uma grande valia para alunos portadores de deficiências (alunos autistas) ou sugerir questões tendo em conta o progresso.

Os projetos MEGUA e SIACUA inserem-se na nova linha de investigação MATEAS-MATEMÁTICA: ENSINO E AVALIAÇÃO NO (ENSINO) SUPERIOR [12] da unidade de investigação CIDMA - Centro de Investigação em Matemática e Aplicações [5].

Capítulo 7

Conclusão

Este trabalho teve como objetivo conhecer e caracterizar a aplicação *web* denominada SIACUA, já implementada e em fase de experimentação, aberta a futuras melhorias, no Departamento de Matemática da universidade de Aveiro, bem como, todo o estudo dos conceitos subjacentes a ela.

Começámos por abordar a nova conceção de ensino a par da evolução tecnológica, em que, em consequência dessa evolução foram criados os sistemas tutores inteligentes. Sendo, portanto, que, o elemento-chave de um sistema tutor inteligente é a capacidade de adaptar-se ao aluno, a componente-chave deste sistema é o chamado modelo estudante, onde a informação do estudante, como, registo de dados do aluno, as suas respostas às questões e o seu padrão de comportamento, é armazenado, durante a sua interação com o sistema.

Quanto à inferência, a partir dos referidos dados do aluno, do seu estado de conhecimento, será baseada no cálculo de probabilidades à posteriori. Estas são afetadas pelas evidências dadas pelo aluno, mas também pelas probabilidades definidas à priori, pelo professor, e portanto, com alguma incerteza e imprecisão, dado o grande número de fatores que podem influenciá-la, tanto a nível de dados de aluno, como a nível dos dados definidos empiricamente pelo professor.

Foi, então, exposto de forma global, não aprofundada, esta ferramenta, em que o con-

ceito básico de redes bayesianas é conhecido como um modelo gráfico que representa uma distribuição de probabilidade conjunta. O objetivo básico da teoria da probabilidade é proporcionar uma explicação coerente de como a crença num dado conhecimento pode ser alterada na presença de informações parciais ou incertas. A função da sua representação gráfica, através de grafos, é mostrar as dependências causais entre as variáveis de maneira clara. Estas variáveis, por sua vez, formam em conjunto com estas relações de causa e efeito, um modelo matemático que representa um domínio real específico, no nosso caso, representa o domínio de conhecimentos numa disciplina.

Para o domínio em estudo nesta dissertação, apresentou-se a forma como foi definido o modelo de redes bayesianas, apresentando-se a sua estrutura, as suas probabilidades condicionadas associadas e, dado que estas crescem exponencialmente com o número de variáveis, foi explicada a forma de como simplificar a sua especificação. Com a utilização do teorema de Bayes, a ocorrência de independência condicional entre variáveis aleatórias que descrevem os dados simplifica os cálculos e também reduz consideravelmente o número de probabilidades condicionadas que precisam ser especificadas. Foi também apresentado a simplificação dos parâmetros a especificar dada uma evidência pelo aluno, ou seja, as respostas às questões, utilizando, para isso, quatro novos parâmetros e uma função da TRI.

A principal dificuldade encontrada ao longo deste trabalho foi a falta de experiência e conhecimentos anteriores sobre redes bayesianas, para o qual foi necessário um encargo extra na exploração dos conteúdos da literatura de RBs existente.

Uma evolução natural do sistema apresentado será a atualização dos diversos parâmetros com base em toda a informação recolhida pelo próprio sistema quando este é utilizado pelos alunos, em vez de usar os parâmetros estimados pelos professores. Para isso podem ser usadas técnicas de *data mining*.

Outra evolução natural consiste na sugestão ao aluno de questões ou temas com base no estudo do seu conhecimento, recolhido pelo modelo bayesiano, podendo para isso usar-se diagramas de inferência.

Referências

- [1] D. M. Cardoso, J. Symański e M. Rostami. *Matemática Discreta*. Escolar Editora, Lisboa, 2008.
- [2] P. Carvalho, J.P. Cruz, L. Descalço, P. Oliveira, D. Seabra. *Messias - a Multidisciplinary E-learning Self-Study InterActive System*. Research Report, Universidade de Aveiro, 2014.
- [3] E. Castillo, J. Guitérrez e A. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, New York, 1997.
- [4] G. Castillo, E. Millán, L. Descalço, P. Oliveira, S. Diogo. *Using Bayesian networks to improve knowledge assessment*. Computers & Education 60(1): 436-447, 2013.
- [5] CIDMA - Centro de Investigação em Matemática e Aplicações, disponível em <http://cidma.mat.ua.pt>
- [6] M. J. Druzdzel. *SMILE: Structural Modeling, Inference, and Learning Engine and GeNIe: A development environment for graphical decision-theoretic models (Intelligent Systems Demonstration)*. Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99), pages 902-903, AAAI Press/The MIT Press, Menlo Park, CA, 1999.
- [7] *Genie & Smile*, Decision Systems Laboratory of the University of Pittsburgh, disponível em <http://genie.sis.pitt.edu>.
- [8] R. C. Guimarães e J. A. Cabral. *Estatística*. McGraw-Hill, Lisboa, 1997.

- [9] F. Jensen e T. Nielsen. *Bayesian Networks and Decision Graphs*. Springer, New York, 2007.
- [10] T. Koski e J. Noble. *Bayesian Networks. An Introduction*. John Wiley and Sons, United Kingdom, 2009.
- [11] M. Martins, C. Monteiro, J. Viana e M. Turkman. *Probabilidades e Combinatória*. Ministério da Educação, Lisboa, 1999.
- [12] *MATEAS-MATEMÁTICA: Ensino e avaliação no (ensino) superior*, disponível em mateas.wikidot.com.
- [13] *Mathematics Exercise Generator, Universidade de Aveiro (Megua)*, disponível em <http://cms.ua.pt/megua/>.
- [14] *MEGUA package for parameterized exercises in Sage Mathematics*, Portuguese Website (<http://cms.ua.pt/megua/>), 2013.
- [15] E. Millán, J.-L. Pérez-de-la-Cruz. *A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation. User Model. User-Adapt. Interact.* 12(2-3): 281-330, 2002.
- [16] E. Millán, T. Lobada, J.-L. Pérez-de-la-Cruz. *Bayesian networks for student model engineering*. *Computers and Education* 55(4): 1663-1683, 2010.
- [17] E. Millán. Thesis: *Bayesian system for student modeling*, 2000.
- [18] B. Murteira, M. Antunes. *Probabilidades e Estatística, vol.1*. Escolar Editora, Lisboa, 2012.
- [19] F. S. Perotto. Thesis: *Modelagem do Conhecimento, Sistemas Especialistas e o Projeto SEAMED*.
- [20] *PmatE (Projeto Matemática Ensino)*, disponível em <http://pmate.ua.pt>.
- [21] E. Reis, P. Melo, R. Andrade e T. Calapez. *Estatística Aplicada*. Edições Sílabo, Lisboa, 1996.

- [22] S.Russell e P. Norving. *Artificial Intelligence: A Modern Approach*. Prentice Hall, New Jersey, 1995.
- [23] *Siacua (Interactive Computer Learning System, University of Aveiro)*, disponível em <http://siacua.web.ua.pt>.
- [24] W. A. Stein et al. *Sage Mathematics Software (Version 4.6.1)*, The Sage Development Team., <http://www.sagemath.org>, 2011.
- [25] *Portal Action*, disponível em <http://www.portalaction.com.br/content/28-distribui%C3%A7%C3%A3o-condicional-caso-discreto>.